



UNIVERSIDAD CARLOS III DE MADRID
Departamento de Teoría de la Señal y Comunicaciones

PROYECTO FIN DE CARRERA

ANÁLISIS DE RECONOCEDORES DE CARAS BASADOS EN VECTORES COMUNES DISCRIMINATIVOS

Autor:

Mercedes Cuevas del Pino

Tutor:

Emilio Parrado Hernández

Leganés, Noviembre de 2009

Título: ANÁLISIS DE RECONOCEDORES DE CARAS BASADOS EN
VECTORES COMUNES DISCRIMINATIVOS

Autor: Mercedes Cuevas del Pino

Director: Emilio Parrado Hernández

EL TRIBUNAL

Presidente: Jose Miguel Leiva

Vocal: Susana Patón

Secretario: Daniel García Doñoro

Realizado el acto de defensa y lectura del Proyecto Fin de Carrera el día 23 de Noviembre de 2009 en Leganés, en la Escuela Politécnica Superior de la Universidad Carlos III de Madrid, acuerda otorgarle la CALIFICACIÓN de

VOCAL

SECRETARIO

PRESIDENTE

Agradecimientos

En primer lugar, quiero dar las gracias a mi tutor, Emilio Parrado, por su orientación y apoyo durante mi proyecto. Este trabajo no hubiera sido posible sin su enseñanza, así como las cualidades humanas y la cantidad de tiempo que me dedicó.

Quiero dar las gracias a mi familia por su continuo apoyo en todos los aspectos, en especial a mis padres porque ellos siempre saben orientarme, y sus consejos siempre me han ayudado. Y por supuesto, a mis hermanos, por vuestra confianza en mí.

A todos mis profesores, desde el colegio hasta la universidad, por todo lo que he aprendido gracias a vosotros.

A todos mis compañeros y amigos de la universidad; Dani, Pablo, Isa, María, David, Cristo, Fran... porque sin todos vosotros, vuestros resúmenes, vuestros consejos, vuestra compañía en las prácticas y las duras épocas de exámenes, seguro que no estaba escribiendo estas líneas.

A todos mis amigos de siempre, Helena, Laura, Belén, Raquel,... que siempre estáis ahí.

A Juan Pablo, por supuesto, porque tu aliento y apoyo incondicional durante toda la carrera han sido muy importantes para mí.

Y a todos aquellos que olvido nombrar.

Gracias.

Resumen

El reconocimiento facial puede ser definido como la identificación de individuos a partir de imágenes de sus caras usando una base de datos de caras almacenada y etiquetada con la identidad de los diferentes sujetos. Esta tarea es compleja y puede descomponerse en pequeños pasos de detección de caras en un fondo de imagen recargado, localización de estas caras seguido de extracción de características de las regiones de la cara y, finalmente, reconocimiento y verificación [2]. Éste es un problema difícil ya que hay numerosos factores como pose 3D, expresión facial, peinado, maquillaje, etc., que influyen en la apariencia de las características faciales del individuo. Además de esta variedad de factores, iluminación, fondo, y cambios de escala hacen esta tarea mucho más exigente. Condiciones problemáticas adicionales incluyen ruido, oclusiones, y otros muchos factores.

En este Proyecto Fin de Carrera se va a evaluar y comparar la robustez de dos métodos de reconocimiento facial, el método de Vectores Comunes Discriminativos, **DCV** (*Discriminative Common Vectors*) y el método Vectores Comunes Discriminativos basados en Kernels, **KDCV** (*Discriminative Common Vectors with Kernels*) en tres variantes: con kernel polinómico de grado dos, con kernel polinómico de grado 3 y con kernel gaussiano. En estudios anteriores, ha sido demostrada la superioridad de estos métodos en términos de reconocimiento, eficiencia y estabilidad numérica. El DCV se comporta favorablemente frente a otros métodos lineales, y el KDCV frente a otros enfoques kernel. El análisis se llevará a cabo previa programación de los algoritmos con la herramienta “matlab”, y será realizado en torno a las

dificultades o variabilidades que nos podamos encontrar en las imágenes, como ruido, oclusiones, iluminación lateral y reescalado, así como en las bases de datos, como el número de muestras de entrenamiento o el número de individuos en la base de datos.

Por lo tanto, el trabajo consistirá en el estudio de los algoritmos así como de las bases de datos que vamos a utilizar, la programación de los métodos y la evaluación de su robustez frente a factores como ruido, reescalado, oclusiones, cambios en la iluminación, número de muestras de entrenamiento y número individuos en la base de datos.

Índice general

<u>1. INTRODUCCIÓN Y OBJETIVOS</u>	<u>15</u>
1.1. INTRODUCCIÓN	15
1.2. MOTIVACIÓN.....	17
1.3. APLICACIONES	20
1.4. DIFICULTADES.....	23
1.4.1. ILUMINACIÓN	23
1.4.2. POSE	24
1.4.3. EXPRESIÓN FACIAL	25
1.4.4. OCLUSIONES PARCIALES	26
1.4.5. OTROS TIPOS DE VARIACIONES	26
1.5. SISTEMAS DE RECONOCIMIENTO FACIAL	27
1.5.1. TAREAS DEL RECONOCIMIENTO FACIAL	27
1.6. EL PROCESO DE RECONOCIMIENTO FACIAL.....	30
1.7. HISTORIA DEL RECONOCIMIENTO FACIAL	34
1.7.1. TÉCNICAS DE APRENDIZAJE MÁQUINA PARA RECONOCIMIENTO.....	36
1.8. OBJETIVOS Y ESQUEMA DE MEMORIA	50
<u>2. MÉTODOS DE CLASIFICACIÓN BASADOS EN VECTORES COMUNES DISCRIMINATIVOS.....</u>	<u>53</u>

2.1. INTRODUCCIÓN	54
2.2. VECTORES COMUNES DISCRIMINATIVOS (DCV)	59
2.3. VECTORES COMUNES DISCRIMINATIVOS BASADOS EN NÚCLEO (<i>KERNEL DCV</i>)	66
2.4. DIFERENCIAS ENTRE LOS MÉTODOS DCV LINEAL Y KERNEL DCV	72
 <u>3. BASES DE DATOS PÚBLICAS PARA RECONOCIMIENTO FACIAL.....</u>	 74
3.1. BASE DE DATOS DE CARAS YALE.....	78
3.2. BASES DE DATOS DE CARAS AR	81
3.3. BASE DE DATOS DE CARAS AT&T AND OLIVETTI.....	84
3.4. BASE DE DATOS DE CARAS GEORGIA TECH	86
3.5. BASE DE DATOS DE CARAS IMM FRONTAL	88
3.6. RESUMEN DE LAS PRINCIPALES BASES DE DATOS UTILIZADAS	89
3.7. ALGUNAS CONSIDERACIONES DE LAS BASES DE DATOS EN RECONOCIMIENTO FACIAL	91
 <u>4. EVALUACIÓN DE EXPERIMENTOS.....</u>	 95
4.1. VERIFICACIÓN DE LA IMPLEMENTACIÓN DE LOS MÉTODOS DCV Y KDCV	97
4.1.1. VERIFICACIÓN DE LA IMPLEMENTACIÓN DEL MÉTODO DCV	97
4.1.2. VERIFICACIÓN DE LA IMPLEMENTACIÓN DE LOS MÉTODOS KDCV: CON KERNEL; POLINÓMICO DE SEGUNDO GRADO, POLINÓMICO DE TERCER GRADO Y GAUSSIANO	100
4.2. ROBUSTEZ DE LOS MÉTODOS DCV Y KDCV FRENTE A LA VARIACIÓN DEL NÚMERO DE INDIVIDUOS EN LA BASE DE DATOS	103
4.3. ROBUSTEZ DE LOS MÉTODOS DCV Y KDCV FRENTE A LA VARIACIÓN DEL NÚMERO DE MUESTRAS	106
4.4. ROBUSTEZ DE LOS MÉTODOS DCV Y KDCV FRENTE A CAMBIOS EN LA ILUMINACIÓN	116
4.5. ROBUSTEZ DE LOS MÉTODOS DCV Y KDCV FRENTE A OCLUSIONES PARCIALES	119
4.6. ROBUSTEZ DE LOS MÉTODOS DCV Y KDCV FRENTE AL RUIDO	121
4.6.1. ROBUSTEZ DE LOS MÉTODOS DCV Y KDCV FRENTE AL RUIDO GAUSSIANO...	123
4.6.2. ROBUSTEZ DE LOS MÉTODOS DCV Y KDCV FRENTE AL RUIDO SAL Y PIMIENTA	131

4.7. ROBUSTEZ DE LOS MÉTODOS DCV Y KDCV FRENTE A LA REDUCCIÓN DE LA RESOLUCIÓN	140
<u>5. CONCLUSIONES</u>	<u>146</u>
5.1. COMPARACIÓN DEL COMPORTAMIENTO DE LOS MÉTODOS	153
<u>REFERENCIAS</u>	<u>155</u>

Capítulo:

1. Introducción y objetivos

1.1. Introducción

El reconocimiento de imágenes faciales, también denominado reconocimiento de caras, permite determinar la identidad de una persona al comparar una imagen de su cara con imágenes de referencia almacenadas en una base de datos, en la que también se almacena la identidad de las personas asociadas a cada imagen de referencia. Esta comparación se realiza analizando características estructurales presentes en las caras. Una declaración general del problema puede ser formulada como: Dadas unas imágenes (de vídeo o no) de una escena [1], identificar o verificar la identidad de una o más personas en la escena usando un depósito de bases de datos de caras. Suele ser útil tener disponible información secundaria como por ejemplo, raza, edad, sexo, expresión facial o locución [2]. La solución del problema supone una segmentación de escenas aglomeradas, detección de caras, extracción de características del rostro y

reconocimiento o verificación. En los problemas de identificación, la entrada al sistema es una cara desconocida y el sistema devuelve la identidad determinada de una base de datos de individuos conocidos, mientras que en los problemas de verificación, el sistema necesita confirmar o rechazar la identidad solicitada de la cara de entrada.

En reconocimiento facial han sido muchas las técnicas que se han desarrollado. Nosotros escogemos la que se basa en Vectores Comunes Discriminativos; el método lineal y los basados en núcleos (*kernels*) (**DCV** y **KDCV**) ya que se han obtenido resultados en investigaciones anteriores que prueban que, en primer lugar, el método DVC es superior a los otros métodos lineales en términos de precisión de reconocimiento, eficiencia y estabilidad numérica. Y en segundo lugar, con el método kernel DVC se ha comprobado que, en muchas situaciones, el comportamiento generalizado de este método se compara favorablemente con otros enfoques kernel. Estos enfoques extraen las propiedades comunes de los individuos en el conjunto de entrenamiento mediante la eliminación de las diferencias en cada clase.

1.2. Motivación

El reconocimiento de rostros es una tarea tan común para los seres humanos, que el individuo ni siquiera advierte el gran número de veces que la realiza al día. Se han realizado investigaciones en reconocimiento facial automático desde la década de los 60, pero ha sido recientemente cuando esta área ha llamado la atención de la comunidad científica. Muchos análisis de caras y técnicas de modelado facial han progresado significativamente en la última década. Sin embargo, la fiabilidad de los sistemas de reconocimiento facial todavía plantea un gran desafío para la comunidad científica.

Aunque los seres humanos son expertos en el reconocimiento facial, todavía no se entiende cómo se realiza este reconocimiento. Durante los últimos 20 años neurocientíficos, psicofísicos e ingenieros han llevado a cabo numerosas investigaciones sobre varios aspectos del reconocimiento facial por humanos y máquinas. Han estado investigando si el reconocimiento de la cara se hace de manera integral o por análisis de rasgos locales, es decir, el reconocimiento de la cara se hace mirando a la cara en su conjunto o mirando rasgos faciales locales independientemente [4,5]. Psicofísicos y neurocientíficos han estado orientados a ideas como: La singularidad de los rostros, si el reconocimiento es global o se hace mediante análisis de rasgos locales; el análisis y el uso de expresiones faciales para reconocimiento, cómo los niños perciben las caras, organización de la memoria para las caras, incapacidad de reconocer con precisión rostros invertidos; existencia de una neurona “abuela”, el papel del hemisferio derecho del cerebro en la percepción de caras; y la incapacidad de reconocer caras debido a condiciones como la prosopagnosia. Algunas de las teorías expuestas para explicar resultados experimentales son contradictorias. Muchas de las hipótesis y teorías hechas por investigadores en estas disciplinas se basan en pequeños conjuntos de imágenes. No obstante, muchos de los hallazgos han tenido consecuencias importantes para los ingenieros que diseñan algoritmos y sistemas para reconocimiento máquina de rostros humanos.

Algunas de las cuestiones de interés potencial para los diseñadores son:

- **¿Es el reconocimiento facial un proceso dedicado? [15]:** Existen evidencias que proceden de tres fuentes. A) Las caras son más fácilmente recordadas por los humanos que otros objetos. B) Los pacientes de prosopagnosia son incapaces de reconocer caras familiares, pero normalmente no tienen otra profunda amnesia. Reconocen a la gente por sus voces, su color de pelo, la vestimenta, etc. Aunque pueden percibir ojos, nariz, boca, pelo, etc., son incapaces de unir todas estas características para realizar la identificación. C) Los recién nacidos parecen preferir mirar a los estímulos en movimiento que tienen patrones como los de la cara en comparación con aquellos que no los tienen o tienen características confusas.
- **¿Es la percepción de caras el resultado de una integración o de análisis de características?:** Ambas, información holística y de características, son cruciales para la percepción y el reconocimiento de caras. Si las características dominantes están presentes, puede que no se usen las descripciones integrales. Por ejemplo, en estudios de recuerdo de caras, los seres humanos rápidamente se centran en las características extrañas, tales como grandes orejas, una nariz torcida, unos ojos mirando fijamente, etc.
- **Ranking de importancia de los rasgos faciales:** Para la apreciación y el recordatorio de caras se han establecido como rasgos importantes, el cabello, el óvalo facial, los ojos y la boca (no necesariamente en este orden). Muchos estudios demuestran que la nariz juega un papel insignificante; puede ser debido al hecho de que todos esos estudios han sido hechos usando imágenes frontales. Otro resultado de varios estudios es que la parte superior de la cara es más útil en reconocimiento que la parte inferior. También se ha estudiado el papel de atributos estéticos como la belleza, atractivo y/o agrado, con la conclusión de que caras más atractivas obtienen tasas de reconocimiento mayores, después las menos atractivas y por último las caras más comunes.

- **El papel del análisis de frecuencia espacial:** Los más recientes estudios [16, 17] determinan que la información en baja frecuencia juega un papel dominante en el reconocimiento facial. Dependiendo de la tarea específica, los componentes en baja, media y alta frecuencia pueden desempeñar papeles diferentes. Por ejemplo, la tarea de determinación de sexo se logra con éxito usando únicamente componentes en baja frecuencia, mientras que la tarea de identificación requiere el uso de componentes en alta frecuencia. Los componentes en baja frecuencia contribuyen a la descripción global, y los componentes en alta frecuencia a los detalles más finos, requeridos en tareas de identificación.
- **El papel de la raza o el sexo:** Los seres humanos reconocen mejor a las personas de su propia raza que a personas de otras razas. Esto puede ser debido al hecho de que los seres humanos pueden que codifiquen una cara “promedio” con atributos “promedios”, cuyas características son diferentes para razas diferentes, haciendo más difícil el reconocimiento de caras de razas diferentes. Goldstein [18] da dos posibles razones de las discrepancias: Una razón psicosocial, en el que los pobres resultados de la identificación son producto de los efectos de los prejuicios, la falta de familiaridad, o una variedad de otras razones interpersonales; y otra razón psicofísica, que hace frente a la pérdida de detalles faciales debido a los diferentes colores de piel, o diferencias relacionadas con la raza en la variabilidad de rasgos faciales. En el estudio [19] se ha mostrado que en una población japonesa, la mayoría de los rasgos faciales de las mujeres son más heterogéneos que los rasgos de los hombres. También se ha encontrado que las caras de las mujeres de raza blanca son ligeramente más variables que las de los hombres, pero que la variación total es pequeña.

1.3. Aplicaciones

La biometría, que se basa en la autenticación de los aspectos intrínsecos de un ser humano, aparece como una alternativa viable a los enfoques más tradicionales (tales como contraseñas o códigos PIN). Entre las más antiguas técnicas biométricas se encuentra el reconocimiento de huellas dactilares. Esta técnica fue utilizada en China ya en el año 700 dC para la certificación oficial de los contratos. Más tarde, en la mitad del siglo XIX, fue utilizada para la identificación de personas en Europa. Una técnica biométrica desarrollada actualmente es el reconocimiento del iris. Esta técnica se utiliza ahora en lugar de la identificación a través del pasaporte para los viajeros frecuentes en algunos aeropuertos del Reino Unido, Canadá y los Países Bajos. Así como para el control de acceso de los empleados a las zonas restringidas de los aeropuertos de Canadá y en el aeropuerto JFK de Nueva York. Estas técnicas suponen un inconveniente debido a la necesidad de interacción con el individuo que va a ser identificado o autenticado. El reconocimiento facial tiene varias ventajas sobre otras técnicas biométricas, es natural, no intrusiva y fácil de usar. Es por ello por lo que esta técnica ha atraído un interés cada vez mayor de la comunidad científica en los últimos años.

El mayor interés en los sistemas de reconocimiento facial automático, adquirido en distintos entornos de la comunidad científica, se debe en gran parte al aumento de las preocupaciones del público en cuanto a seguridad, especialmente debido a las numerosas situaciones de terror en torno al mundo después del 11 de septiembre de 2001. El interés por el reconocimiento facial automático y la cantidad de aplicaciones, es muy probable que aumenten más aún en el futuro. Esto podría ser debido al aumento de la penetración de las tecnologías, tales como cámaras digitales e Internet, y debido a una mayor demanda de los diferentes regímenes de seguridad. Además, la falsificación de tarjetas de identidad o la intrusión en zonas físicas o virtuales mediante descifrado de contraseñas alfanuméricas aparecen con frecuencia en los medios de comunicación. Estos problemas de la sociedad moderna han desencadenado una verdadera necesidad

de mecanismos de control fiables, de fácil uso y aceptación general para la identificación y verificación de las personas.

Las posibles aplicaciones de la autenticación de caras se ocupan principalmente del control de acceso, por ejemplo, restringir el acceso físico a un edificio, tales como edificios de empresas, zonas seguras de un aeropuerto, una casa, etc. en lugar de abrir una puerta mediante una clave o un código, la respectiva persona se comunicaría con un identificador, por ejemplo, su nombre, y presentaría su cara frente a una cámara. El sistema de autenticación de caras entonces comprobaría la identidad de la persona y concedería o denegaría el acceso. Este principio podría aplicarse igualmente para el acceso a los sistemas, cajeros automáticos, teléfonos móviles, sitios de Internet, etc. donde se tendría que presentar la cara ante una cámara en lugar de introducir un número de identificación o contraseña.

Evidentemente, también puede llevarse a cabo el control de acceso de una persona tan sólo presentando su rostro a la cámara sin declarar su identidad. Un sistema reconocimiento de la identidad de una persona puede, mucho mejor, ser empleado para controlar más específicamente los derechos de las personas respectivas almacenados en su base de datos. Por ejemplo, los padres pueden permitir a sus hijos ver sólo ciertos programas de televisión o sitios web, mientras la televisión o la computadora están reconociendo automáticamente las personas que se sitúan enfrente.

La vídeo vigilancia es otra aplicación del reconocimiento facial relacionada con la seguridad. El objetivo aquí es reconocer a los sospechosos o delincuentes utilizando cámaras de vídeo instaladas en lugares públicos, tales como bancos o aeropuertos, a fin de aumentar la seguridad general de estos lugares. En este contexto, la base de datos de sospechosos a reconocer es a menudo muy grande y las imágenes captadas por la cámara son de baja calidad, lo que hace que la tarea sea más difícil.

Finalmente, existen otras muchas aplicaciones posibles, distintas de las aplicaciones orientadas a la seguridad (control de acceso o sistemas de verificación, sistemas de vigilancia,...), como de entretenimiento e interacción personalizada hombre - máquina. En el campo de la avanzada Interacción Hombre-Máquina (*Human-Computer Interaction*, HCI), un ejemplo es la facilitación de la interacción de personas discapacitadas con ordenadores u otras máquinas o el reconocimiento automático de las

expresiones faciales a fin de detectar la reacción de la persona sentada delante de la cámara (por ejemplo, sonreír, reír, bostezar dormir).

Con la amplia propagación de las cámaras digitales en los últimos años el número de imágenes digitales almacenadas en los servidores y ordenadores personales ha crecido con rapidez. En consecuencia, existe una necesidad cada vez mayor de sistemas que automáticamente clasifiquen y anoten esta cantidad enorme de imágenes con el fin para permitir una búsqueda efectiva.

1.4. Dificultades

Existen algunas propiedades inherentes a las caras así como a la forma en que las imágenes son capturadas, que hacen del procesamiento automático de imágenes faciales una tarea complicada. En el caso de reconocimiento facial, nos encontramos con el problema de la varianza dentro de la clase, es decir, las variaciones de la cara de la misma persona debido a la iluminación, la pose, etc., que son, con frecuencia, mayores que la variación entre clases; es decir, variaciones de la apariencia facial de diferentes personas. Este hecho reduce la tasa de reconocimiento. En muchas aplicaciones de análisis facial, la variación de la apariencia, resultante de estas circunstancias puede considerarse también como ruido y hace que sea más difícil extraer la información deseada, es decir, la identidad de la persona. Además, reduce el comportamiento general de los respectivos sistemas.

A continuación, resumiremos las dificultades más importantes encontradas en aplicaciones comunes en el mundo real.

1.4.1. Iluminación

Cambios en la iluminación pueden producir variaciones considerables en la apariencia de los rostros y por tanto, en las imágenes de los rostros. Son dos las principales fuentes de luz que proporcionan la iluminación general: la luz ambiente y un punto de luz (o luz dirigida.) La primera es, de alguna forma, más fácil de manejar ya que solo afecta al brillo general de la imagen resultante. Sin embargo, la segunda es más difícil de analizar ya que las imágenes tomadas bajo direcciones variantes de fuentes de luz siguen una función altamente no lineal. Además, la cara puede producir sombras en sí misma. La figura 1.1 muestra el impacto de las diferencias de iluminación en los rostros.



Figura 1.1: Ejemplo de cara bajo un punto fijo de observación y con variación de iluminación.

Se han propuesto muchas investigaciones para tratar con este problema. Algunos métodos de detección o reconocimiento facial tratan de ser invariantes a los cambios de iluminación mediante el modelado implícito de la misma o la extracción de características invariantes. Otros estudios proponen un paso de procesamiento a parte, una especie de normalización para reducir los efectos de los cambios de iluminación.

1.4.2. Pose

La variación de la posición de la cabeza, en otras palabras, al ángulo de visión desde el que se toma la imagen de la cara, es otra dificultad que impacta fuertemente en el comportamiento de los métodos automáticos de análisis facial. Por este motivo, muchas aplicaciones se limitan a imágenes frontales o, en otro caso, ejecutan un procesamiento de pose específica que requiere una estimación precedente de la pose, como en aproximaciones de reconocimiento facial en múltiples vistas.

Si la rotación de la cara coincide con el plano de la imagen, la pose puede ser normalizada estimando el ángulo de rotación y girado la imagen de tal manera que la cara esté en una posición vertical. Este tipo de normalización es parte de un procedimiento llamado *Alineamiento facial* o *declaración facial*.

En la Figura 1.2 podemos observar un ejemplo de imágenes de un sujeto sometido a variaciones de pose cuando mantenemos el resto de factores, como expresión o iluminación, constantes.



Figura 1.2: Ejemplo de cara bajo iluminación fija y variación de pose

1.4.3. Expresión facial

La apariencia de una cara con diferentes expresiones, varía considerablemente (ver Figura 1.3). Dependiendo de la aplicación, esto puede ser más o menos importante. Por ejemplo, para sistemas de control de acceso normalmente se les solicita a los individuos que muestren una expresión neutra. Por lo tanto, la invarianza de la expresión facial no puede ser un problema en este caso. En caso contrario, en un sistema de indexación de vídeo o imagen, por ejemplo, podría ser más importante que las personas se muestren en situaciones cotidianas y puede que hablando, sonriendo, etc.

En general, la boca es objeto de la mayor variación. Una persona en una imagen puede tener la boca abierta o cerrada, puede estar hablando, sonriendo o incluso haciendo muecas.

Los ojos y las cejas cambian en el individuo al modificar la expresión facial, por ejemplo, cuando una persona pestañea, duerme o abre ampliamente sus ojos.



Figura 1.3: Ejemplo de cara bajo iluminación y pose fijas pero con variación en la expresión facial.

1.4.4. Oclusiones parciales

Las oclusiones parciales ocurren poco frecuentemente en imágenes en el mundo real. Pueden ser causadas porque se tape parte de la cara con las manos, por ejemplo, la boca, o con el pelo, gafas u otros objetos o personas.

En la mayoría de los casos, sin embargo, la cara oculta parte de ella misma. Por ejemplo, en una vista desde un lado, se tapa el otro lado de la cara. También puede ocurrir que parte de las mejillas se oculten por la nariz o un ojo puede ser cubierto por su órbita por ejemplo.

1.4.5. Otros tipos de variaciones

Las variaciones en la apariencia también pueden ser causadas por variación en el maquillaje, variación en el corte de pelo, y la presencia de vello facial (barba, bigote, etc).

La variación de edad es también un factor influyente en el comportamiento de métodos de análisis facial. Este es el caso, por ejemplo, en reconocimiento facial, cuando las imágenes de referencia han sido tomadas años antes de hacer el reconocimiento.

Finalmente, hay variaciones también a lo largo de las identidades de los sujetos, tales como raza, color de piel, o más generalmente, origen étnico. Las respectivas diferencias en la apariencia de las imágenes de rostros pueden causar dificultades en aplicaciones como detección de características faciales o reconocimiento de sexo.

1.5. Sistemas de Reconocimiento Facial

1.5.1. Tareas del Reconocimiento Facial

Las tres tareas primarias de reconocimiento facial son:

- Verificación (autenticación) - ¿Soy quién digo ser? (búsqueda uno entre uno)
- Identificación (reconocimiento) - ¿Quién soy? (búsqueda uno entre muchos)
- Lista de vigilancia - ¿Me estás buscando a mi? (búsqueda uno entre varios)

Se han aplicado diferentes esquemas para evaluar las tareas descritas anteriormente. Qué esquema ha de usarse depende de la naturaleza de la aplicación.

VERIFICACIÓN

La tarea de verificación está dirigida a aplicaciones que requieren la interacción del usuario reivindicando su identidad, es decir, aplicaciones de acceso. El test de verificación se lleva a cabo mediante la división de personas en dos grupos:

- **Cientes**, personas que intentan acceder utilizando su propia identidad.
- **Impostores**, personas que tratan de acceder utilizando una identidad falsa, es decir, una identidad conocida por el sistema, pero que no pertenece a ellas.

El porcentaje de impostores que acceden se reporta como la Tasa de Falsa Aceptación (*False Acceptance Rate*, FAR) y el porcentaje de acceso rechazado a clientes se reporta como la Tasa de Falso Rechazo (*False Rejection Rate*, FRR) para un determinado umbral. Un ejemplo de esto se puede ver en la Figura 1.4

IDENTIFICACIÓN

La tarea de identificación es destinada, en su mayoría, a aplicaciones que no requieran la interacción del usuario, es decir, aplicaciones de vigilancia. El test de identificación trabaja desde la asunción de que todas las caras en el test son personas conocidas. El porcentaje de identificaciones correctas es entonces reportado como la Tasa de Identificación Correcta (CIR) o el porcentaje de falsa identificación es reportado como Tasa de Falsa Identificación (FIR).

LISTA DE VIGILANCIA

La tarea de Lista de Vigilancia es una generalización de la tarea de identificación que incluye gente desconocida. El test de Lista de Vigilancia es como el test de identificación reportado en CIR o FIR, pero puede tener FAR y FRR asociadas a él para describir la sensibilidad de la lista de vigilancia, lo que significa con qué frecuencia es clasificada una persona desconocida como persona en la lista de vigilancia (FAR).

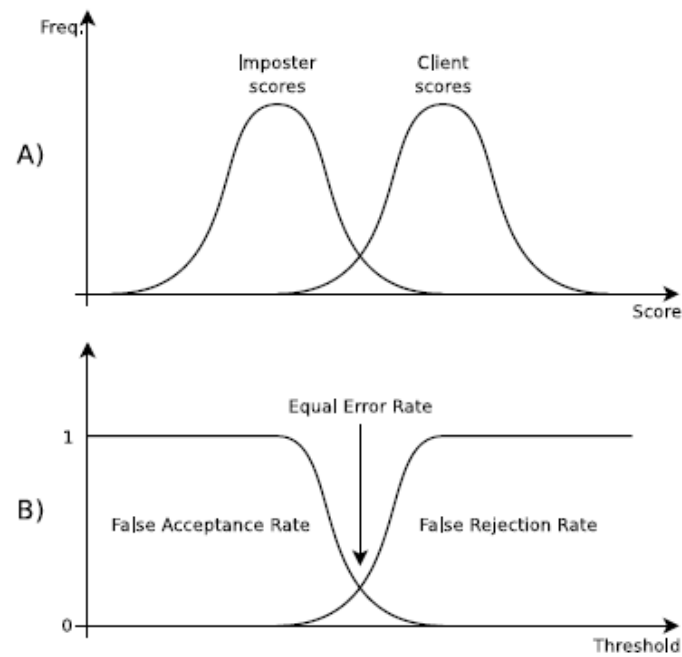


FIGURA 1.4: Relación de la Tasa de Falsa Aceptación (FAR). Tasa de Falso Rechazo (FRR) con la distribución de clientes, impostores en un esquema de verificación. A) Muestra las poblaciones de clientes e impostores en términos de puntuación (alta puntuación significa alta probabilidad de pertenecer a la población de clientes). B) Las asociadas FAR y FRR, la Tasa de Igual Error (EER) está donde las curvas FAR y FRR se encuentran y da el valor umbral para la mejor separabilidad de las clases impostor y cliente.

1.6. El proceso de Reconocimiento Facial

El reconocimiento facial es una tarea de reconocimiento visual de muestras. Se reconocen las caras humanas tridimensionales, que son susceptibles de variaciones como iluminación, pose, expresiones, etc. Este reconocimiento puede llevarse a cabo con varias fuentes de entrada de datos como: Una única imagen 2D, imágenes 2D en estéreo (dos o más imágenes 2D) o escaneos láser 3D.

Recientemente, debido a las aplicaciones militares, comerciales y solicitudes legales, ha habido un gran interés en el reconocimiento automático de caras en fotogramas e imágenes de vídeo. Esta investigación abarca varias disciplinas como procesamiento de imágenes, reconocimiento de patrones, visión por computador y aprendizaje máquina. Los datos proceden de una extensa variedad de fuentes. Un grupo de fuentes son aquellas imágenes de formato relativamente controlado, como pasaportes, tarjetas de crédito, carnés de conducir, y fotos de archivos policiales. Una clase más variable de imágenes de aplicación incluyen detecciones en tiempo real y reconocimiento de caras en imágenes de vídeo-vigilancia, lo que presenta limitaciones adicionales en términos de velocidad y requisitos de procesamiento [3]. Para construir estas diferentes aplicaciones, se necesitan técnicas y métodos de reconocimiento facial automático que sean precisos y robustos. Sin embargo, estas técnicas y métodos en la actualidad no están disponibles o están únicamente disponibles en escenarios de alta complejidad y coste.

Además, en breve, las cámaras de 3D de Tiempo de Vuelo (*Time Of Flight*, TOF) serán lo suficientemente precisas para ser utilizadas también. La dimensión de estas fuentes puede ser incrementada en uno, por la introducción de una dimensión tiempo. Una imagen fija con una dimensión de tiempo es una secuencia de vídeo. La ventaja es que la identificación de una persona puede ser determinada, con más precisión, de una

secuencia de vídeo que de una imagen fija ya que la identidad de una persona puede que no cambie en dos fotos tomadas a continuación, en una secuencia de vídeo.

En este proyecto, nos centraremos en la transformación de dos dimensiones a nivel de gris de las imágenes ya que ésta es la forma más extendida de imágenes digitales y, por tanto, permite la propuesta de criterios que se aplicarán en la más amplia y genérica forma. Sin embargo, muchas técnicas descritas en este trabajo podrían extenderse también a imágenes de color, de datos en 3D o datos multimodales.

Esta compleja tarea del reconocimiento puede descomponerse en pequeños pasos, detección de caras en un fondo de imagen recargado, localización de estas caras seguido de extracción de características de las regiones de la cara y, finalmente, reconocimiento y verificación. En este proceso además se encuentra la tarea previa de adquisición de la imagen, en ocasiones, con preprocesado. Podemos resumirlos, como indica la Figura 1.5 en cuatro pasos; detección de la cara (localización), preprocesado de la imagen (alineamiento/normalización facial, corrección de luz, etc.), extracción de características y coincidencia de características.

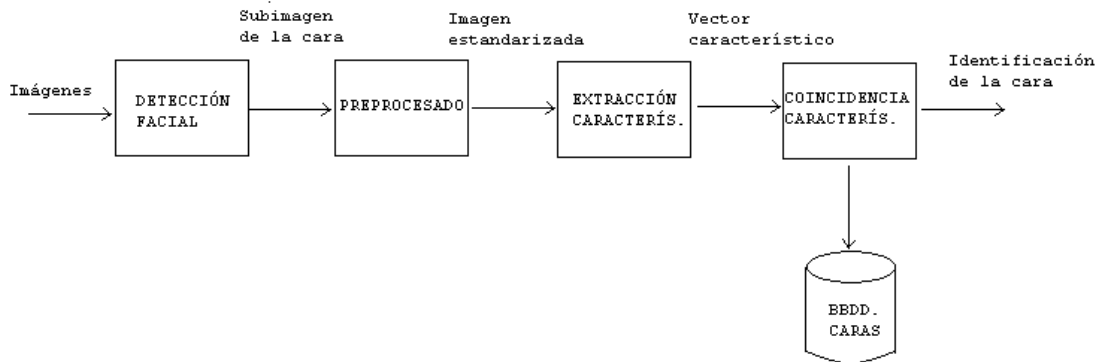


Figura 1.5: Los cuatro pasos generales del reconocimiento facial.

DETECCIÓN DE CARAS

La detección o localización facial denota el problema de encontrar una o varias subregiones en imágenes, cada subregión conteniendo una cara, es decir, determinar la posición y escala de las respectivas regiones faciales y en ocasiones también su ángulo de rotación con el plano de la imagen. Los algoritmos de detección de rostros, en su

mayoría representan los resultados como las aristas de las cajas que cubren las regiones de la imagen que se supone contienen las caras, es decir, un cuadro delimitador por la cara. Se ha de tener en cuenta que la mayor parte de las tareas de análisis facial, por ejemplo reconocimiento facial, reconocimiento de la expresión facial, detección de las características faciales, requiere de la localización de caras como paso inicial.

La detección facial puede implicar un paso de pre-procesado para preparar la imagen de entrada y realizar una especie de normalización, por ejemplo, convertir la imagen a escala de gris, ecualizar el histograma de la imagen o de las regiones de la imagen. Normalmente, este pre-procesado se considera como parte del algoritmo de detección facial.

Adicionalmente, puede haber un paso de pos-procesado para una mayor normalización del resultado del procedimiento de detección de rostros, por ejemplo, los cuadros delimitadores. Operaciones posibles son la normalización de la iluminación, la normalización o la estimación de la pose, la alineación de la cara, esto logra la alineación global de los cuadros delimitadores. Nos referimos a estos procedimientos como *normalización facial*. La normalización facial es un paso importante en muchas aplicaciones de análisis de cara porque elimina la información visual irrelevante para la tarea correspondiente. Por ejemplo, en el reconocimiento de caras, las diferentes condiciones de iluminación pueden causar que las imágenes de una misma persona aparezcan muy diferentes, mientras que dos imágenes de diferentes personas bajo la misma condición de iluminación podrían ser muy similares. Así, sin una normalización de la iluminación previa, la clasificación general puede ser bastante difícil.

En el caso en el que la entrada sea vídeo, puede ser una ventaja rastrear la cara entre múltiples fotogramas. Los métodos usados para detección incluyen: Plantillas de Forma (*Shape Templates*), Redes Neuronales y Modelos de Apariencia Activa (*Active Appearance Models*, AAM.)

La detección de caras se lleva a cabo mediante la detección de características faciales, que consiste en la localización de puntos característicos en imágenes faciales, como ojos, nariz, boca, etc. Los algoritmos de detección de características difieren en varios aspectos. Primero, pueden ser parte de un método de detección facial donde los detectores de características escanean una gran imagen que puede o no contener caras.

Después, dada una localización con la presencia de todas o la mayoría de las características faciales se presupone la presencia de una cara en esa posición particular. Muchos algoritmos de detección de características faciales, sin embargo, operan en imágenes o regiones que contienen únicamente la cara que ha sido localizada y extraída de antemano por un algoritmo de detección facial. Otro punto donde los algoritmos difieren es el tipo de imagen en la que trabajan. Esto implica la resolución de la imagen, así como las condiciones externas tales como la iluminación, el ruido, el fondo, además de la gama de posibles posiciones de cabeza y expresiones faciales. Otro aspecto importante es el propósito o la aplicación de destino de los respectivos algoritmos de detección de características ya que esto podría imponer más o menos limitaciones de precisión, así como el tipo y número de características de detectar.

PREPROCESADO

Mediante el paso de preprocesado se normaliza la gruesa detección facial, de manera que se pueda conseguir una robusta extracción característica. Dependiendo de la aplicación, el preprocesado de las imágenes de caras incluye: Alineación (translación, rotación, escalado) y normalización/correlación de luz.

EXTRACCIÓN DE CARACTERÍSTICAS

Este paso trata de extraer un conjunto compacto de características discriminatorias geométricas/fotométricas de la cara. Los métodos para extracción característica incluyen: PCA, FLDA y Proyecciones (LPP).

COINCIDENCIA DE CARACTERÍSTICAS

La coincidencia de características es el actual proceso de reconocimiento. El vector característico obtenido de la extracción característica se hace coincidir con las imágenes faciales de clases (personas) almacenadas en la base de datos.

1.7. Historia del Reconocimiento Facial

El camino más intuitivo para llevar a cabo el reconocimiento facial es mirar las características principales de la cara y compararlas con las mismas características de otras caras. Algunos de los primeros estudios en reconocimiento facial fueron dados por Darwin [6] y Galton [7]. El trabajo de Darwin incluye un análisis de las diferentes expresiones debido a estados emocionales distintos, donde como Galton, estudió perfiles faciales. Sin embargo, los primeros intentos reales de desarrollar un sistema de reconocimiento facial semiautomático comenzaron a finales de los años 60 y principios de los 70, y estaban basados en información geométrica. Se pusieron, en las fotografías, marcas en las principales características faciales, tales como ojos, orejas, nariz, y comisuras de la boca. Se calcularon distancias relativas y ángulos a un punto común de referencia y se compararon con los datos de referencia.

Una aproximación más consistente para hacer reconocimiento facial fue hecha por Fischler y colaboradores [8] (1973) y más tarde por Yuille y colaboradores [9] (1992). Este estudio medía los componentes faciales usando plantillas de un único componente y los proyectaba en una plantilla global.

Resumiendo, la mayoría de las técnicas desarrolladas durante las primeras etapas del reconocimiento facial se centraron en la detección automática de las características faciales del individuo. Las mayores ventajas de estos métodos basados en características geométricas son la insensibilidad a la iluminación y la comprensión intuitiva de las características extraídas. Sin embargo, aún hoy la detección de características faciales y las técnicas de medición no son suficientemente fiables para el reconocimiento de caras basado en características geométricas y estas características, por sí solas, no son suficientes para el reconocimiento facial.

Por estos motivos, esta técnica fue gradualmente abandonada y se hicieron esfuerzos en la investigación de técnicas basadas en color, que han proporcionado mejores resultados. Las técnicas basadas en color alinean un conjunto de caras diferentes para obtener una correspondencia entre intensidades de píxeles y se puede usar un clasificador de vecino más próximo para clasificar nuevas caras cuando la nueva imagen es primero alineada al conjunto de imágenes ya alineadas. Mediante la aparición de la técnica de Autocaras (*Eigenfaces*) [10], un estudio de aprendizaje estadístico, se mejoró notablemente este burdo método. En lugar de comparar directamente la intensidad de píxel de las diferentes imágenes, primero se redujo la dimensión de las intensidades mediante el Análisis de Componente Principales (*Principal Component Analysis*, PCA) en la técnica de Eigenfaces. Eigenfaces es un componente base de otros esquemas de reconocimiento usados hoy en día. Una de las técnicas actuales es la de Caras de Fisher (*Fisherfaces*). Esta técnica es ampliamente utilizada [11, 12]. Combina Eigenfaces con Análisis de Discriminante Lineal de Fisher (*Fisher Linear Discriminant Analysis*, FLDA) para obtener una mejor separación de los rostros de los individuos. En Fisherfaces, la dimensión de los vectores de intensidad se reduce mediante PCA y luego se aplica FLDA para obtener una proyección óptima para la separación de las caras de diferentes personas. PCA y FLDA se describen con más detalles en el capítulo 2.

Después del desarrollo de la técnica Fisherface, se han propuesto otros muchos métodos relacionados. Estas nuevas técnicas ayudan a proporcionar una mejor proyección para la separación de las caras de diferentes personas. Intentan reforzar la robustez para hacer frente a las variaciones de iluminación o de pose. Entre estas técnicas encontramos Caras de Fisher basadas en Núcleo (*Kernel Fisherface*) [13], Caras de Laplace (*Laplacianfaces*) [14] o el método de Vectores Comunes Discriminativos (**DCV**) con su versión basada en núcleo (**Kernel DCV**), que serán los métodos objeto de estudio en este Proyecto.

1.7.1. Técnicas de aprendizaje máquina para reconocimiento

A continuación, resumiremos algunas de las aproximaciones más importantes en aprendizaje máquina con el objeto del reconocimiento facial. Las técnicas de aprendizaje máquina aprenden automáticamente, del conjunto de muestras, cómo clasificar nuevas instancias del mismo tipo de datos. Una propiedad importante de estas aproximaciones, y que debe ser maximizada, es la capacidad de generalización, es decir, la habilidad de clasificar con éxito datos desconocidos y posiblemente establecer reglas genéricas o funciones. Normalmente, distinguimos entre dos tipos de aprendizaje:

- **Aprendizaje supervisado:** Tenemos disponibles un conjunto de entrenamiento y la correspondiente salida deseada de la función para poder aprender. Así, durante el entrenamiento, el algoritmo presenta iterativamente ejemplos al sistema y adapta sus parámetros de acuerdo a la distancia entre la salida producida y la deseada.
- **Aprendizaje no supervisado:** La estructura subyacente del conjunto de datos, es decir, la salida deseada, es desconocida y será determinada por el algoritmo de entrenamiento. Por ejemplo, para un método de clasificación esto significa que la información de la clase no está disponible y tenemos que aproximarnos a ella agrupando las muestras de entrenamiento y usando algunas distancias de medida, una técnica llamada “agrupamiento” (*clustering*).
- **Aprendizaje reforzado:** Aquí, la salida exacta de la función a aprender es desconocida, y el entrenamiento consiste en un ajuste de parámetro basado únicamente en dos conceptos, premio y castigo. Si el sistema no se comporta bien (suficientemente) es “penalizado” y los parámetros se adaptan como corresponda. En caso contrario es premiado, es decir, tienen lugar algunos refuerzos positivos.

La mayoría de los algoritmos descritos a continuación son supervisados, pero se emplean con fines muy diferentes: algunos de ellos son utilizados para extraer características de los datos de entrada, algunos son utilizados para clasificar las características extraídas y otros realizan ambas tareas.

El contexto de aplicación varía también en gran medida, es decir, algunos de los enfoques pueden ser utilizados para detección de características y/u objetos, algunos únicamente para reconocimiento y otros para ambos. Además, en muchos sistemas se usa una combinación de varias de las técnicas que a continuación se describen. En este sentido, podrían ser considerados como una especie de bloques para sistemas efectivos de reconocimiento y detección.

1.7.1.1. METODOS DE PROYECCIÓN ESTADÍSTICA

Para ser capaces de analizar imágenes automáticamente, con frecuencia éstas se redimensionan para tener un cierto ancho w y alto h . Después las filas o columnas de las respectivas imágenes se concatenan para formar un vector de dimensión $d = w \times h$. El espacio del vector resultante se llama *espacio de imagen*, y de aquí en adelante lo denotamos con I .

En las tareas de procesamiento de señales, frecuentemente hay mucha redundancia en las respectivas imágenes/vectores porque, en primer lugar, las imágenes de algunas clases son bastante similares y, en segundo lugar, los píxeles próximos en una imagen están altamente correlados.

De esta manera, parece obvio representar la imagen en un formato más compacto, es decir, proyectar los vectores en un subespacio S de I , en lo que consiste el método de proyección estadística. En la literatura, el término *reducción de la dimensión* o *selección característica* se emplean con frecuencia en el contexto de estas técnicas. Estos métodos ayudan en el cálculo de S que, en general, es de menor dimensión que I , lo que hace que

los vectores de imágenes sean estadísticamente menos correlados. Hay principalmente dos grupos de proyecciones: *lineales* y *no lineales*.

Las técnicas de proyección lineal transforman el vector de una imagen $x = (x_1, \dots, x_n)^T$ de dimensión d en un vector $s = (s_1, \dots, s_k)^T$ de dimensión k , mediante una matriz de transformación lineal W , de dimensión $k \times d$:

$$s = W^T \cdot x$$

En general, se eliminan aquellos vectores base que se supone contienen la información menos importante para una aplicación dada utilizando un criterio predefinido. Así, la dimensión k del espacio resultante S se puede escoger después de calcular los vectores base que generan el subespacio.

Los métodos de proyección más comunes son el PCA y el LDA, que ya hemos nombrado anteriormente.

Las aproximaciones no lineales son aplicadas cuando las proyecciones lineales no son suficientes para representar los datos de manera que permitan la extracción de características discriminatorias. Este es el caso para distribuciones más complejas donde meros hiper-planos no fallen al tratar de separar las clases a distinguir. Como la mayoría de los enfoques son iterativos, requieren una elección *a priori* de la dimensión k del subespacio resultante S .

1.7.1.1.1. Análisis de Componentes Principales

También conocido como Transformada de Karhunen-Loève (KLT) o Transformada Hotelling [20], es una proyección lineal ortogonal dentro del subespacio donde la primera dimensión (o eje) corresponde con la dirección de I que tiene la mayor varianza, la segunda a la dirección con la segunda mayor varianza y así sucesivamente.

Dado cierto número N de vectores de entrada $\{x_1, x_2, \dots, x_N\}$ ($x_i \in \mathbb{R}^n$) que se supone tienen una distribución multi-normal y están centrados, es decir, $1/N \cdot \sum_{i=1}^N (x_i) = 0$, los correspondientes vectores de proyección son:

$$S_i = W^T \cdot x_i \quad i \in 1, \dots, N$$

donde $s_i \in \mathbb{R}^k$. Ahora dado Σ la matriz de covarianza de los vectores de entrada:

$$\Sigma = \frac{1}{N} \sum_{i=1}^N x_i x_i^T$$

Entonces la matriz de covarianza de los vectores proyectados s_i se define como

$$\Sigma' = W^T \Sigma W$$

Finalmente, se propone la matriz de proyección para maximizar la varianza de los vectores proyectados. Así,

$$W = \arg \max_{\tilde{W}} | \tilde{W}^T \Sigma \tilde{W} |$$

Las k columnas de W , es decir, los vectores base de S , se llaman componentes principales y representan los autovectores correspondientes a los mayores autovalores de la matriz de covarianza Σ .

Una característica importante de PCA es que si $k < d$ el error de reconstrucción e en términos de distancia Euclídea es mínimo,

$$e = \frac{1}{N} \sum_{i=1}^N |x_i - W_{s_i}|$$

Así, los k primeros autovectores forman un subespacio que representa de manera óptima el espacio de entrada I .

Aparte de la compresión de la imagen y coincidencia de plantillas, se aplica PCA para tareas de clasificación, por ejemplo, los enfoques *Eigenfaces* [21] en reconocimiento facial. Aquí, los vectores de proyección s_i son el objeto de clasificación. Para este propósito las N imágenes de entrada son cada una asociada con una etiqueta de clase y usada para construir un clasificador. El clasificador más simple sería el clasificador de vecino más cercano usando una medida de distancia Euclídea.

Para resumir, PCA calcula el subespacio lineal ortogonal que tiene sus ejes orientados con las direcciones de varianzas mayores. Esto representa entonces los datos de entrada óptimamente. Sin embargo, en un contexto de clasificación no se garantiza

que en el subespacio calculado mediante PCA se mejore la separación de los datos. En este aspecto, LDA es más apropiado.

1.7.1.1.2. Análisis de Discriminante Lineal

Fisher [22] introdujo el LDA en 1936 pero lo generalizó después en el llamado Discriminante Lineal de Fisher (*Fisher's Linear Discriminant*, FLD). Además de la mejor representación de los datos de entrada, también trata la separabilidad en el subespacio proyectado con respecto a las diferentes clases.

Dado $\Omega = \{x_1, \dots, x_N\}$ un conjunto de entrenamiento dividido en c clases denotadas Ω_i ($i \in 1, \dots, c$). Nosotros ahora buscamos el subespacio S que maximiza la variabilidad inter-clase mientras se maximiza la variabilidad intra-clase, así mejoramos la separabilidad de las respectivas clases. Para conseguir esto maximizamos el llamado criterio de Fisher [22, 23]:

$$J(W) = \frac{|W^T \Sigma_b W|}{|W^T \Sigma_w W|} \quad \text{Así,}$$

$$W = \arg \max_{\tilde{W}} \frac{|\tilde{W}^T \Sigma_b \tilde{W}|}{|\tilde{W}^T \Sigma_w \tilde{W}|}$$

donde $\Sigma_w = \frac{1}{N} \sum_{j=1}^c \sum_{x_i \in \Omega_j} (x_i - \bar{x}_j)(x_i - \bar{x}_j)^T$ es la matriz de varianza dentro de la clase y

$\Sigma_b = \frac{1}{N} \sum_{j=1}^c \sum_{x_i \in \Omega_j} N_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})^T$ es la matriz de varianza entre clases. N_j es el

número de muestras en la clase Ω_j , \bar{x}_j son las respectivas muestras y \bar{x} es la media de todos los datos.

1.7.1.1.3. Otros métodos de proyección

Existen otras muchas otras técnicas de proyección propuestas en la literatura y que pueden aplicarse a detección y reconocimiento.

Por ejemplo, Análisis de Componentes Independientes (ICA) [24] es una técnica que se usa frecuentemente para separación ciega de fuentes [25], es decir, para encontrar las diferentes fuentes independientes de las que una señal está compuesta. ICA busca un subespacio lineal donde los datos no solamente estén incorrelados, sino que sean estadísticamente independientes. El modelo es:

$$\mathbf{x} = \mathbf{A}^T \cdot \mathbf{s},$$

donde \mathbf{x} es el dato observado, \mathbf{s} son las fuentes independientes y \mathbf{A} es la matriz mezcla. ICA consiste en la optimización (que puede estar basada en diferentes criterios) de una función objetivo, llamada función contraste. La función contraste debe asegurar que los datos proyectados sean independientes y no gaussianos, Notar que ICA no reduce la dimensionalidad de los datos de entrada. De hecho, se emplea con frecuencia junto con PCA u otras técnicas de reducción de la dimensionalidad. Existen numerosas implementaciones de ICA, por ejemplo, INFOMAX [24], JADE [26] o FastICA [27].

Yang y colaboradores [28] introdujeron el llamado PCA bidimensional, que no requiere que la imagen de entrada sea transformada en un vector de una sola dimensión previamente. De hecho, se estima directamente una matriz de covarianza generalizada usando las matrices de las imágenes.

Visani y colaboradores [29] presentaron una aproximación similar basada en LDA: el bidimensional orientado a LDA. El procedimiento es análogo al método 2D-PCA donde la proyección se hace directamente en las matrices de las imágenes, por columnas o por filas. Para obtener la matriz de proyección se minimiza un criterio de Fisher generalizado.

Los métodos de proyección presentados son técnicas de proyección lineal. Sin embargo, en algunos casos las diferentes clases no pueden separarse correctamente en un subespacio lineal. Entonces, los métodos de proyección *no lineal* pueden ayudar a mejorar las tasas de clasificación. La mayoría de los métodos de proyección lineal

pueden convertirse en no lineales mediante la proyección de los datos de entrada en un espacio de mayor dimensión donde las clases sean más propensas a ser linealmente separables. Esto significa que el hiperplano de separación en este subespacio representa un subespacio no lineal del espacio de vector de entrada. Afortunadamente, no es necesario describir explícitamente este espacio de mayor dimensión y la respectiva función de proyección si encontramos una función *núcleo* (*kernel*) que implementa un producto escalar en su espacio de vector. La función kernel permite evaluar un producto en el espacio vectorial de destino y puede usarse para construir versiones no lineales de las técnicas de proyección descritas anteriormente, como por ejemplo PCA, LDA o ICA.

Los enfoques de proyección que hemos repasado anteriormente en esta introducción pueden ser aplicados, principalmente, a cualquier tipo de datos para realizar un análisis estadístico de las respectivas muestras. También se puede clasificar como aproximación de proyección estadística una técnica llamada Modelo de Apariencia Activa (AAM) [30], aunque está mucho más especializada en modelar imágenes u objetos deformables bajo variación de condiciones externas. Así, al contrario que los métodos PCA o LDA, donde las imágenes de entrada son tratadas como un vector “estático”, ahora se tienen en cuenta pequeñas deformaciones.

1.7.1.2. MODELOS DE APARIENCIA ACTIVA

Los AAM, fueron introducidos por Cottes y colaboradores [30] como una extensión de los Modelos de Forma Activa (ASM) [31], y representan un enfoque que describe estadísticamente no sólo la textura de un objeto, sino también la forma. Dada una nueva imagen de la clase de objetos a analizar, la idea es interpretar el objeto sintetizando una imagen del mismo mientras se va aproximando tan bien como sea posible su apariencia en la imagen real. Esto ha sido principalmente aplicado a los problemas de análisis facial. El modelado de la forma de las caras parece útil en la mayoría de las aplicaciones de análisis facial donde las imágenes de los rostros son susceptibles a cambios en la pose y en la expresión facial.

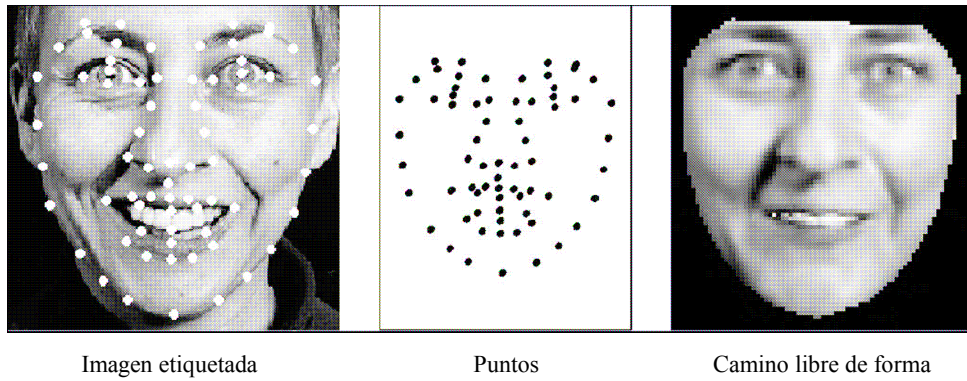


Figura 1.6: Modelos de Apariencia Activa: muestra de entrenamiento marcada y el correspondiente camino libre de forma

1.7.1.3. MODEOS OCULTOS DE MARKOV

Este método probabilístico que representa, por ejemplo, una imagen, en una *secuencia* de observaciones. Fue introducido por Rabiner y colaboradores [32, 33] y se usa para modelar el aspecto secuencial de los datos. En el contexto de procesamiento de señales, por ejemplo, se aplican reconocimiento de voz modelando las secuencias temporales de los estados y observaciones, por ejemplo, fonemas. Una imagen puede verse también como una secuencia de observaciones, por ejemplo, subregiones de la imagen, y aquí las imágenes tienen que ser también linealizadas en una estructura unidimensional.

1.7.1.4. IMPULSO ADAPTATIVO

El Impulso Adaptativo (*Adaboost*) es una técnica de clasificación introducida por Freund y Schapire [34]. La idea básica aquí es combinar varios clasificadores “débiles” en un clasificador “fuerte”, donde los clasificadores débiles se comportan sólo ligeramente mejor que el azar aleatorio.

El principio del algoritmo es aprender una función de decisión global aleatoria mediante la suma y el entrenamiento iterativo de clasificadores débiles, por ejemplo, redes Wavelets, mientras se centra en más y más ejemplos difíciles. Esto se ha aplicado a muchos problemas de clasificación y ha llegado a usarse ampliamente como técnica de aprendizaje máquina debido a su simplicidad y comportamiento en términos de tasas de clasificación y tiempo de cómputo.

1.7.1.5. MÁQUINA DE VECTOR SOPORTE

La técnica de clasificación llamada Máquina de Vector Soporte (SVM) [35, 36, 37] está basada en la Minimización de Riesgo Estructural (SRM) formulada por Vapnik y colaboradores en [38]. Una de las ideas básicas de esta teoría es que la tasa de error de test, o riesgo estructural $R(\alpha)$, es limitada superiormente por la suma de las tasas de error de entrenamiento, riesgo empírico R_{emp} y un término adicional denominado confianza, VC, que depende de la llamada dimensión VC h de la función de clasificación.

1.7.1.5.1. Máquinas de Vector Soporte Lineales

Vapnik mostró la función de decisión para un hiperplano lineal:

$$f(\mathbf{x}) = \text{sign}((\mathbf{w} \cdot \mathbf{x}) + b),$$

donde la dimensión VC se determina mediante la norma de \mathbf{w} .

Dado $\{(x_i, y_i), \dots, ((x_l, y_l))\}$, con $x_i \in \mathbb{R}^n$, $y_i \in \{-1, +1\}$ el conjunto de entrenamiento. Para un conjunto de entrenamiento linealmente separable tenemos:

$$y_i (\mathbf{x}_i \cdot \mathbf{w} + b) - 1 \geq 0 \text{ para todo } i = 1, \dots, l. \text{ (condición 1.1)}$$

El margen entre los puntos positivos y negativos se define por dos hiperplanos $\mathbf{x} \cdot \mathbf{w} + b = \pm 1$ donde el término de arriba es actualmente cero. La figura 1.7 se ilustra esto. Además, no se encuentran puntos entre estos hiperplanos y el ancho del margen es $2 / \|\mathbf{w}\|$. El algoritmo de vector soporte ahora intenta maximizar el margen mediante la

minimización de $\|w\|$, lo que supone que es una solución óptima, es decir, donde la generalización es máxima. Una vez obtenido el margen máximo, los datos que están sobre uno de los hiperplanos de separación, es decir, para los que la ecuación anterior son cero, se llaman *vectores soporte* (ilustrados como dobles círculos en la figura 1.7).

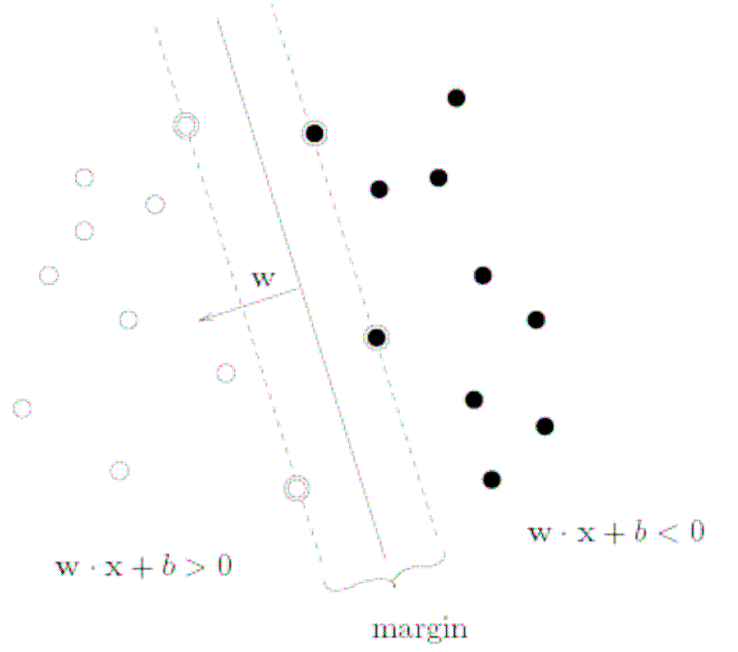


Figura 1.7: Ilustración gráfica del SVM lineal

Para simplificar el cálculo, el problema se formula en marco de trabajo de Lagrange. Se trata de maximizar los Lagrangianos:

$$L_D = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{ij} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad ,$$

sujeto a $w = \sum_{i=1}^l \alpha_i y_i x_i$, $\sum_{i=1}^l \alpha_i y_i = 0$ y $\alpha_i \geq 0 \quad \forall i = 1 \dots l$, donde α_i ($i = 1, \dots, l$) son los multiplicadores de Lagrange que han de determinarse. Además, la solución de α_i y la condición 1.1 implican el valor de b . Todos los α_i son cero excepto aquellos que corresponden a los vectores soporte. Finalmente, se pueden simplificar nuevas muestras usando la función de decisión descrita anteriormente.

En muchos casos, sin embargo, los datos de entrenamiento no se pueden separar completamente porque son valores extremos (“*outliers*”). Entonces, simplemente se debe relajar la condición 1.1 introduciendo una constante $\xi_i > 0$ de la siguiente forma:

$$y_i (x_i \cdot w + b) \geq (1 - \xi_i) \text{ para todo } i = 1, \dots, l; \text{ donde ahora } 0 \leq \alpha_i \leq \xi_i.$$

1.7.1.6. BOLSAS DE FIRMAS LOCALES

En oposición a las SVMs, es una técnica de clasificación muy general. La aproximación *Bag of Local Signatures* (BOLS) fue introducida recientemente por Csurka y colaboradores [39] para problemas de clasificación de imágenes, particularmente detección de objetos y reconocimiento. Fue motivada por el enfoque “bolsa de palabras” para la categorización de texto que simplemente cuenta el número de palabras clave predefinidas en un documento para clasificarlo en una o varias categorías de texto.

En el primer paso del método, se detectan n puntos salientes $\mathbf{p}_i = (x_i, y_i)$ de la imagen, usando un algoritmo de detección de punto de interés. La pequeña imagen alrededor de cada punto detectado se representa mediante algunos descriptores, produciendo una firma local \mathbf{s}_i para cada punto saliente. En el paso siguiente, se clasifican las firmas extraídas aplicando cualquier método de cuantificación de vector. Con este fin, se calcula un diccionario de k firmas representativas \mathbf{d}_i del conjunto de entrenamiento usando un algoritmos de agrupamiento. Después, se obtiene una bolsa de firmas locales, \mathbf{v}_i , es decir, entradas del diccionario, representando la apariencia del objeto en la imagen. Se calculan histogramas $h(\mathbf{I})$ de \mathbf{v}_i simplemente contando el número de ocurrencias de las respectivas firmas. Finalmente se clasifican los histogramas usando una medida de distancia de histogramas.

La Figura 1.6 muestra el procedimiento general para generar una representación BOLS. Una ventaja de este método comparado con métodos de proyección estadística es, por ejemplo, su robustez frente a oclusiones parciales y cambios de pose del objeto a reconocer.

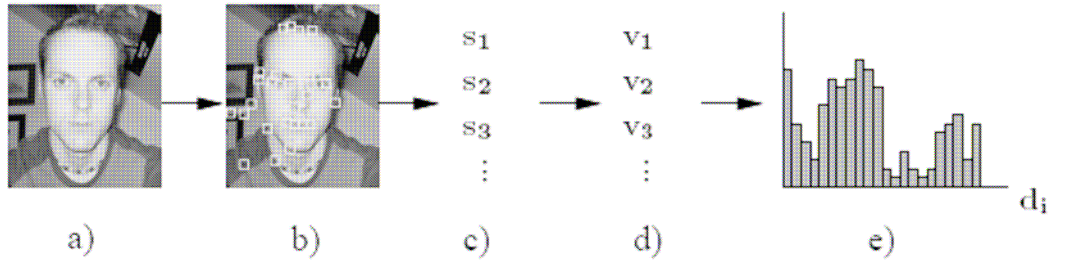


Figura 1.6: Procedimiento de creación de histogramas con aproximación BOLS: a) imagen de entrada I , b) puntos salientes detectados, c) firmas locales extraídas, d) vectores cuantificados (entradas en el diccionario), e) histograma h .

1.7.1.7. REDES NEURONALES

Las Redes Neuronales Artificiales (*Artificial Neural Networks*, ANN), comúnmente abreviadas, “NN”, denotan una técnica de aprendizaje inspirada en el cerebro humano y su capacidad de tratar tareas complejas mediante la actuación de neuronas interconectadas, realizando cada una, una operación muy simple. Así mismo, una NN es una estructura entrenable que consiste en un conjunto unidades interconectadas, implementando cada una función muy sencilla, y eventualmente entre todas realizando una función compleja de clasificación o aproximación.

1.7.1.7.1. Perceptrón

El tipo más conocido de la unidad neuronal se llama *Perceptrón* y ha sido introducido por Rosenblatt [40]. Su estructura básica se ilustra en la Figura 1.9. Tiene n entradas y una salida donde la salida es una función simple de la suma de las señales de entrada x ponderada por w y un sesgo adicional b . Así

$$y = \Phi(x \cdot w + b)$$

Con frecuencia, se pone el sesgo dentro del vector de pesos w , tal que $w_0 = b$ y el vector de entrada x se extiende consecuentemente con $x_0 = 1$. Podemos entonces la ecuación anterior como:

$$y = \Phi(x \cdot w)$$

donde Φ es la función paso de Heavyside:

$$\Phi: R \rightarrow R$$

$$\Phi(x) = 1 \quad \text{si } x \geq 0$$

$$\Phi(x) = 0 \quad \text{en caso contrario}$$

El Perceptrón entonces implementa un sencillo clasificador de dos clases donde w es el hiperplano de separación tal que $w \cdot x \geq 0$ para muestras de una clase y $w \cdot x < 0$ para muestras de la otra.

Con el uso de varios perceptrones interconectados podremos formar fronteras de decisión más complejas, mediante la combinación de varios hiperplanos. El tipo más común de estas NNs es el Perceptrón Multicapa, del que hacemos una breve descripción a continuación.

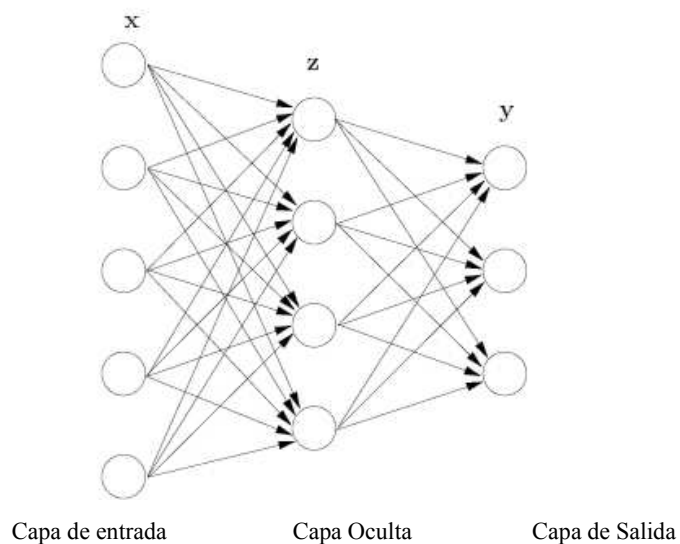
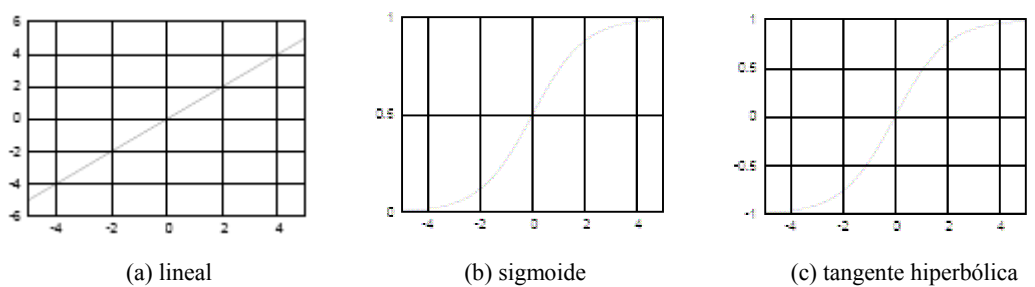
1.7.1.7.2. Perceptrón Multicapa

El Perceptrón Multicapa (MLP) es capaz de aproximar arbitrariamente funciones de decisión complejas. La Figura 1.9 muestra la estructura de un MLP. Tiene una capa de entrada, una o más capas ocultas y una capa de salida de neuronas, donde cada una excepto las neuronas de entrada, implementan un perceptrón como el descrito anteriormente. La activación de las neuronas se propaga a lo largo de las capas desde la capa de entrada hasta la de salida. Si existe una conexión entre cada neurona y cada una de las neuronas de la capa siguiente, la red se denomina *completamente conectada*. Además, la función de activación de las neuronas tiene que ser diferenciable para ajustar los pesos mediante el algoritmo Propagación Hacia Atrás. Las funciones de activación usadas comúnmente son, por ejemplo:

$$\phi(x) = x \quad \text{lineal}$$

$$\phi(x) = \frac{1}{1 + e^{-cx}} \quad (c > 0) \text{ sigma}$$

$$\phi(x) = \frac{1 + e^{-x}}{1 + e^{+x}} \quad \text{tangente hiperbólica}$$


Figura 1.9: Perceptrón Multicapa

Figura 1.10: Diferentes tipos de funciones de activación.

1.8. Objetivos y esquema de memoria

El objetivo principal de este proyecto es evaluar los métodos basados en Vectores Comunes Discriminativos, **DCV lineal** y **kernel DCV**. A su vez, para el método kernel DCV, usaremos 3 tipos de kernel, lo que dará lugar a tres submétodos más. Utilizaremos un kernel polinómico de grado 2, otro de grado 3 y un tercer kernel gaussiano.

Nuestro propósito concreto va a ser comparar la robustez de estos métodos cuando los enfrentamos a diferentes bases de datos con factores de variabilidad diversos. Es decir, haremos un análisis comparativo del comportamiento de estos 4 algoritmos con respecto a las variables: ruido (gaussiano y “sal y pimienta” con diferentes varianzas y densidades respectivamente), condiciones lumínicas, oclusiones, diezmado de las imágenes, número de individuos en la base de datos y número de imágenes por individuo usadas para entrenamiento.

La memoria se compondrá de la siguiente manera: Tras esta introducción, en el capítulo 2 presentaremos los métodos que vamos a evaluar. Haremos una enumeración y breve descripción de los métodos anteriores y relacionados con los métodos DCV hasta ahora desarrollados para clasificación y explicaremos los fundamentos de los métodos elegidos así como la confección de los algoritmos.

En el capítulo 3 haremos lo propio con las bases de datos. Haremos una revisión de las bases de datos que tenemos públicamente disponibles. Es necesaria una mención especial de las bases de datos que hemos escogido para realizar los experimentos y de las características que les hacen interesantes para su estudio. Estas características van a ser el número de individuos, el número de imágenes por individuo de las que se disponen, las variaciones de la posición de los rostros, las diferentes condiciones lumínicas, y la existencia o no de oclusiones.

En el capítulo 4 explicaremos el desarrollo de las diferentes pruebas llevadas a cabo. Expondremos el escenario confeccionado para poder probar la robustez de los métodos basados en Vectores Comunes Discriminativos. Además se presentarán los resultados obtenidos en nuestros experimentos

Por último, en el capítulo 5 haremos una exposición de las conclusiones obtenidas a partir de la sección anterior. Presentaremos un resumen de la robustez del comportamiento de los métodos respecto a las variabilidades introducidas, clasificaremos los métodos en función de su consistencia con respecto a cada factor.

Capítulo:

2. Métodos de clasificación basados en Vectores Comunes Discriminativos

Se han propuesto muchos métodos para el caso de reconocimiento facial [41] en las últimas dos décadas, la mayor parte de las aproximaciones, basadas en apariencia, operan con imágenes procesadas como muestras bidimensionales. En estas aproximaciones, se representa una imagen bidimensional de tamaño $w \times h$ píxeles mediante un vector en un espacio wh -dimensional. Por tanto, cada imagen facial corresponde a un punto en este espacio. Este espacio se denomina “espacio de muestra” o “espacio de imagen”, y su dimensión es típicamente elevada. Sin embargo, como las imágenes de la cara tienen una estructura similar, los vectores de imagen están correlados, y cualquier imagen en el espacio de muestra puede ser representada en un subespacio de menor dimensión sin perder información significativa.

2.1. Introducción

Los algoritmos propuestos para su evaluación, son algoritmos de clasificación que dado el conjunto de entrenamiento compuesto de C clases, donde cada clase contiene N muestras, y dado x_m^i un vector columna d -dimensional ($d=w \cdot h$) que denota a la muestra m -ésima de la clase i -ésima, construyen un clasificador multiclase para otras muestras que denominamos muestras de test. En el conjunto de entrenamiento habrá un total de $M = N \times C$ muestras. Suponemos que $d > M - C$. En este caso, las “matrices de dispersión intra clase” S_W , “matrices de dispersión interclase” S_B , y “matrices de dispersión total” S_T , se definen como:

$$S_W = \sum_{i=1}^C \sum_{m=1}^N (x_m^i - \mu_i)(x_m^i - \mu_i)^T, \quad (1)$$

$$S_B = \sum_{i=1}^C N(\mu_i - \mu)(\mu_i - \mu)^T, \quad (2)$$

$$S_T = \sum_{i=1}^C \sum_{m=1}^N (x_m^i - \mu)(x_m^i - \mu)^T = S_W + S_B, \quad (3)$$

donde μ es la media de todas las muestras, y μ_i es la media de las muestras en la clase i -ésima, definiéndose como:

$$\mu_i = \frac{1}{N} \sum_{m=1}^N x_m^i \quad \text{y} \quad \mu = \frac{1}{M} \sum_{i=1}^C \sum_{m=1}^N x_m^i$$

La matriz S_W es la matriz de dispersión dentro de la clase, es decir la matriz que contiene la suma de todas las varianzas y se obtiene sumando el cuadrado de la resta de cada muestra con la media de las muestras de su clase. La matriz S_B es la matriz de dispersión entre clases y representa la dispersión de la media de cada clase respecto a la media global. Por último, la matriz de dispersión total, S_T , es la suma de las dos

anteriores y se obtiene mediante la suma de la varianza de todas las muestras respecto de la media global. En la Figura 2.1 podemos ver una representación de estos conceptos en el caso bidimensional.

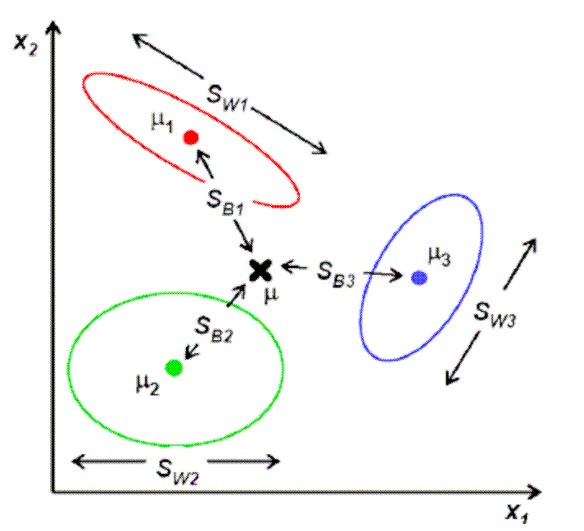


Figura 2.1: Representación de las matrices de dispersión para el caso bidimensional

Se han propuestos numerosos métodos anteriormente a los basados en Vectores Comunes Discriminativos (*Discriminative Common Vectors*, **DCV**). Uno de ellos, el método Autocara (*Fisherface*) [42] trata de encontrar precisamente un subespacio de dimensión menor. La idea de este método que utiliza Análisis de Componentes Principales, (*Principal Component Análisis*, **PCA**) [20], es encontrar el mejor conjunto de direcciones de proyección en el espacio de muestra que maximice la dispersión total a través de todas las imágenes; es decir,

$$J_{PCA}(W_{opt}) = \arg \max_W |W^T S_T W| \quad (4)$$

W en este caso, es la matriz cuyas columnas son los vectores de proyección ortonormal. Las direcciones de proyección se llaman también “autocarar” (*eigenfaces*). Entonces se puede aproximar cualquier imagen en el espacio de muestra mediante combinación lineal de autocaras. Este método es una técnica no supervisada, ya que no considera las clases dentro del conjunto de entrenamiento, es decir, eligiendo un criterio que maximice la dispersión total, esta aproximación trata de modelar variaciones indeseadas dentro de una clase como son las resultantes de diferentes condiciones lumínicas, expresiones faciales y otros factores. Además, como el criterio

no intenta minimizar la variación dentro de la clase, puede que las clases resultantes tengan más solapamiento. Otro método propuesto en [43] y [44], el de Análisis de Discriminante Lineal (*Linear Discriminant Analysis*, **LDA**), supera las limitaciones del método Autocara aplicando el criterio de Discriminante Lineal de Fisher, criterio que intenta maximizar el ratio

$$J_{FLD}(W_{opt}) = \arg \max_W \frac{|W^T S_B W|}{|W^T S_W W|}. \quad \text{CRITERIO FISHER LDA} \quad (5)$$

De esta manera, encontramos las direcciones de proyección que por un lado maximizan la distancia Euclídea entre imágenes de clases diferentes y por otro minimizan la distancia entre imágenes de la misma clase. Esto ocurre cuando W es autovector de $S_W^{-1}S_B$. En tareas de reconocimiento facial este método no puede ser aplicado directamente ya que la dimensión del espacio de muestra es típicamente mayor que el número de muestras en el conjunto de entrenamiento. Como consecuencia, S_W es singular en este caso.

Estos casos, típicamente producidos en reconocimiento facial, en los que la dimensión del espacio de muestra ($d=w \cdot h$) es mayor que el número de muestras en el conjunto de entrenamiento (M), se conocen como casos con “problema de pequeño tamaño de muestra”. Para superar esta situación se propusieron también otros métodos, Tian y colaboradores propusieron el método de la Pseudoinversa [45] que reemplaza S_W^{-1} por su pseudoinversa; el método de Perturbación que suma una matriz de pequeña perturbación, Δ , a S_W para hacerla no singular; el método Descomposición de Rango [46], que autodescompone sucesivamente las matrices S_T y S_B . Sin embargo, estos métodos son costosos computacionalmente ya que las matrices de dispersión son muy grandes, por ejemplo, imágenes de tamaño $256 * 256$ producen matrices de dispersión de tamaño $65536 * 65536$. Se propusieron también otros métodos como Fisherface [47], que utiliza PCA para reducción de la dimensión (con el fin de hacer S_W no singular) para después aplicar LDA, o el método Espacio Nulo [48] en el que todas las muestras de imágenes primero se proyectan en el espacio nulo de S_W y después se aplica PCA para obtener los vectores de proyección óptimos. Utiliza el criterio de Discriminante Lineal de Fisher modificado:

$$J_{MFLD}(W_{opt}) = \arg \max_W \frac{|W^T S_B W|}{|W^T S_T W|}. \text{ CRITERIO MODIFICADO FISHER LDA } (6)$$

El problema que presenta el método Fisherface es al aplicar PCA para la reducción de la dimensión, podemos estar eliminando dimensiones que contienen información significativa, mientras que en el caso de Espacio Nulo no se propuso un algoritmo eficiente para la aplicación de este método en el espacio de muestra original. Además se ha demostrado que cualquier tipo de preprocesamiento que reduzca el espacio de muestra original debe evitarse ya que se reduce también el rendimiento.

Existen otros métodos como PCA+Espacio Nulo [49], en el que primero se aplica PCA para eliminar el espacio nulo de S_T y después se encuentran los vectores de proyección óptima en el espacio restante de menor dimensión utilizando el método de Espacio Nulo. Yang y colaboradores propusieron también una variación pero en ambos casos, aunque se trabaja con el espacio de muestra original, son poco prácticos para las aplicaciones de reconocimiento facial cuando el tamaño del conjunto de entrenamiento es amplio, ya que el coste computacional del entrenamiento llega a hacerse muy grande.

Por último, y anteriormente a nuestros métodos objeto, se propusieron el método LDA Directo [50], que utiliza diagonalización simultánea. En este caso también se elimina el espacio nulo de S_B mediante reducción de la dimensión, de manera que se eliminará también parte del espacio nulo de S_W y puede que se pierda información significativa importante.

En este proyecto, nos proponemos probar la robustez de otros métodos desarrollados, los basados en Vectores Comunes Discriminativos (*Discriminative Common Vectors*, **DCV**). El primero de ellos, es el método **DCV** [51], se basa en una variación del LDA de Fisher para el caso de pequeño tamaño de muestra; y el segundo, el **kernel DCV** [52], basado en “núcleos”, se propuso también para tratar el problema anteriormente mencionado. “¿Cómo debe uno elegir los vectores de proyección óptima para extracción de características en el caso de pequeño tamaño de muestra?” Como respuesta a la anterior pregunta, ambos métodos fueron desarrollados. Tanto **DCV** como **kernel DCV** producen una solución óptima para maximizar el criterio del Discriminante Lineal de Fisher modificado dado anteriormente. Bajo determinadas

condiciones, se garantiza un 100% de la tasa de reconocimiento para las muestras del conjunto de entrenamiento en ambos casos.

2.2. Vectores Comunes Discriminativos (DCV)

Este método se basa en una variación del Análisis de Discriminante Lineal de Fisher para el caso de tamaño pequeño de muestra. Este algoritmo utiliza la matriz de dispersión de las muestras en el conjunto de entrenamiento, S_W , para extraer los vectores comunes discriminatorios que representan a cada persona en el conjunto de entrenamiento de las bases de datos.

Usamos entonces en nuestro estudio comparativo el método DVC, ya que trata con el problema de pequeño tamaño de muestra a la vez que supera las limitaciones de otros métodos propuestos para este mismo fin. De este modo, el método DCV solo puede utilizarse cuando la dimensión del espacio de muestra es mayor que el rango de S_W . La idea de los vectores comunes es extraer las propiedades comunes de las clases en el conjunto de entrenamiento por eliminación de las diferencias de las muestras en cada clase. Se obtiene un vector común para cada clase individual, eliminando todas las componentes que se encuentran en la dirección de los autovectores correspondientes a autovalores distintos de cero de la matriz de dispersión total de su propia clase. Se usan entonces vectores comunes para el reconocimiento. En este caso, en lugar de utilizar la matriz de dispersión total de la propia clase (S_T), utilizamos la matriz de dispersión dentro de la clase de todas las clases (S_W) para obtener los vectores comunes. Para clasificación se usará un nuevo conjunto de vectores obtenidos a partir de los vectores comunes, denominados vectores comunes discriminatorios.

Como hemos introducido anteriormente, dado el conjunto de entrenamiento compuesto de C clases, donde cada clase contiene N muestras, y dado x_m^i un vector columna d -dimensional que denota a la muestra m -ésima de la clase i -ésima. Habrá un total de $M = N \times C$ muestras en el conjunto de entrenamiento. Suponemos que $d > M - C$.

En el caso especial en el que $w^T S_W w = 0$ y $w^T S_B w \neq 0$, para todos los $w \in R^d \setminus \{0\}$, el criterio de Discriminante Lineal de Fisher modificado (ec. 6), alcanza un máximo. Sin embargo, un vector de proyección w , que satisfaga las condiciones anteriores, no maximiza necesariamente la dispersión entre clases. En este caso, se proporciona el siguiente mejor criterio:

$$J(W_{opt}) = \arg \max_{|W^T S_W W|=0} |W^T S_B W| = \arg \max_{|W^T S_T W|=0} |W^T S_T W| \quad (7)$$

Para encontrar el vector de proyección óptimo ω en el espacio nulo de S_W , proyectamos las muestras de caras en el espacio nulo de S_W y obtenemos los vectores de proyección realizando PCA. De hacer esto, se tienen que calcular previamente los vectores que generan el espacio nulo de S_W . Sin embargo, esta tarea es computacionalmente intratable ya que la dimensión de este espacio nulo puede ser muy grande. Un camino más eficiente para llevar a cabo esta tarea es usar el complemento ortogonal del espacio nulo de S_W , que típicamente es un espacio significativamente de menor dimensión. Dado R^d , el espacio de muestra original, V el espacio de rango de S_W , y V^\perp el espacio nulo de S_W ; equivalentemente,

$$V = \text{span}\{\alpha_k \mid S_W \alpha_k \neq 0, k = 1, \dots, r\} \quad (8)$$

$$V^\perp = \text{span}\{\alpha_k \mid S_W \alpha_k = 0, k = r + 1, \dots, d\} \quad (9)$$

donde $r < d$ es el rango de S_W , $\{\alpha_1 \dots \alpha_d\}$ es un conjunto ortonormal, y $\{\alpha_1 \dots \alpha_r\}$ es el conjunto de autovectores ortonormales correspondientes a los autovalores distintos de cero de S_W . Consideramos las matrices $P = [\alpha_1 \dots \alpha_r]$ y $\bar{Q} = [\alpha_{r+1} \dots \alpha_d]$. Ya que $R^d = V \oplus V^\perp$, todas las imágenes de caras $x_m^i \in R^d$ tienen una única descomposición de la forma:

$$x_m^i = y_m^i + z_m^i \quad (10)$$

donde $y_m^i = P x_m^i = Q Q^T x_m^i \in V$, $z_m^i = \bar{Q} \bar{Q}^T x_m^i \in V^\perp$, y P y \bar{Q} son los operadores de proyección ortonormal en V y V^\perp respectivamente. Nuestro objetivo es calcular:

$$z_m^i = x_m^i + y_m^i = x_m^i - Px_m^i \quad (11)$$

Para hacer esto, necesitamos encontrar una base V , que puede ser conseguida por autoanálisis de S_W . En particular, los autovectores normalizados α_K correspondientes a los autovalores distintos de cero de S_W serán una base ortonormal de V . Los autovectores se pueden obtener calculando los autovalores de la matriz más pequeña $A^T A$, definida como $S_W = A^T A$, donde A es una matriz $d \times M$ de la forma:

$$A = [x_1^1 - \mu_1 \dots x_N^1 - \mu_1 \dots x_1^2 - \mu_2 \dots x_N^C - \mu_C], \quad (12)$$

$$\text{siendo } \mu_i = \frac{1}{N} \sum_{m=1}^N x_m^i$$

Dado λ_K y v_K , el autovalor distinto de cero k -ésimo y el correspondiente autovector de $A^T A$, donde $k < M - C$; entonces, $\alpha_K = A \cdot v_K \cdot \lambda_K^{(-1/2)}$ será el autovector que corresponde al autovalor distinto de cero k -ésimo de S_W . La proyección encontrada en V^\perp se logra mediante (11). De esta manera, resulta que obtenemos el mismo y único vector para todas las muestras de la misma clase,

$$\begin{aligned} x_{com}^i &= x_m^i - QQ^T x_m^i = \bar{Q} \bar{Q}^T x_m^i, \\ m &= 1, \dots, N; i = 1, \dots, C, \end{aligned} \quad (13)$$

es decir, el vector de la parte derecha de (13) es independiente del índice m de la muestra. Nos referimos a los vectores x_{com}^i como los vectores comunes. Tras obtener estos vectores comunes, los vectores de proyección óptimos serán aquellos que maximicen la dispersión total de los vectores comunes,

$$J(W_{opt}) = \arg \max_{|W^T S_W W|=0} |W^T S_B W| = \arg \max_{|W^T S_W W|=0} |W^T S_T W| = \arg \max_W |W^T S_{com} W|, \quad (14)$$

donde W es una matriz cuyas columnas son los vectores de proyección óptimos ortonormales w_K , y S_{com} es la matriz de dispersión de los vectores comunes,

$$S_{com} = \sum_{i=1}^C (x_{com}^i - \mu_{com})(x_{com}^i - \mu_{com})^T, \quad i=1, \dots, C, \quad (15)$$

donde μ_{com} , es la media de todos los vectores comunes,

$$\mu_{com} = \frac{1}{C} \sum_{i=1}^C x_{com}^i.$$

En este caso, los vectores de proyección óptimos w_K se pueden encontrar mediante un análisis de S_{com} . En particular, todos los autovectores correspondientes a los autovalores distintos de cero de S_{com} serán los vectores de proyección óptimos. S_{com} es típicamente una matriz grande, $d \times d$, por tanto podemos usar la matriz más pequeña $A_{com}^T A_{com}$, de tamaño $C \times C$, para encontrar los autovalores distintos de cero y los correspondientes autovectores de $S_{com} = A_{com}^T A_{com}$, donde A_{com} es una matriz $d \times C$ de la forma

$$A_{com} = [x_{com}^1 - \mu_{com} \dots x_{com}^C - \mu_{com}]. \quad (16)$$

Habrán $C - 1$ vectores de proyección óptimos, ya que el rango de S_{com} es $C - 1$ si todos los vectores comunes son linealmente independientes. Si dos vectores comunes son iguales, entonces las dos clases que están representadas por ese vector no pueden ser distinguidas. Como los vectores de proyección óptimos w_K pertenecen al espacio nulo de S_W , esto cumple que cuando las muestras de la imagen x_m^i de la clase i -ésima son proyectadas en el espacio lineal generado por los vectores de proyección w_K , el vector característico $\Omega_i = [\langle x_m^i, w_1 \rangle \dots \langle x_m^i, w_{C-1} \rangle]^T$ de los coeficientes de proyección $\langle x_m^i, w_k \rangle$ serán también independientes de la muestra m . Así tenemos,

$$\Omega_i = W^T x_m^i, \quad m=1, \dots, N; \quad i=1, \dots, C. \quad (17)$$

Llamamos a los vectores característicos Ω_i , *vectores comunes discriminatorios* y serán usados para la clasificación de imágenes faciales. El hecho de que Ω_i no dependa del índice m en (17), garantiza el 100% de exactitud en el reconocimiento de muestras en el conjunto de entrenamiento. Esta garantía no ha sido mencionada en relación a otros métodos.

Para reconocer una imagen de test, x_{test} , el vector característico de esta imagen de test se encuentra como,

$$\Omega_{test} = W^T X_{test} \quad (18)$$

Que es comparado posteriormente con el vector común discriminativo Ω_i para cada clase usando la distancia Euclídea. El vector común discriminativo encontrado más

cercano a Ω_{test} es usado para identificar la imagen de test. Ya que Ω_{test} se compara solamente con un solo vector para cada clase, el reconocimiento es muy eficiente en tareas de reconocimiento facial en tiempo real.

A continuación detallamos el algoritmo, con los pasos resumidos que han de llevarse a cabo.

ALGORITMO que usa el rango de S_W

- **Paso 1**

Calcular los autovalores distintos de cero (D) y los correspondientes autovectores (V) de S_W usando la matriz $A^T A$, donde $S_W = A A^T$ y $A = [x_1^1 - \mu_1 \dots x_N^1 - \mu_1 \dots x_1^2 - \mu_2 \dots x_N^2 - \mu_2 \dots x_1^C - \mu_C \dots x_N^C - \mu_C]$.

Formar el conjunto de autovectores correspondientes a autovalores no nulos de S_W , $Q = [\alpha_1 \dots \alpha_r]$, donde r es el rango de S_W . $Q = A \cdot V \cdot D^{(-1/2)}$

- **Paso 2**

Elegir cualquier muestra de cada clase y proyectarla dentro del espacio nulo de S_W para obtener los vectores comunes

$$x_{com}^i = x_m^i - Q Q^T x_m^i, m = 1, \dots, N; i = 1, \dots, C. \quad (19)$$

- **Paso 3**

Calcular los autovectores ω_K de S_{com} , correspondientes a los autovalores distintos de cero, usando la matriz $A_{com}^T A_{com}$, donde $S_{com} = A_{com}^T A_{com}$, y $A_{com} = [x_{com}^1 - \mu_{com} \dots x_{com}^C - \mu_{com}]$. Como mucho hay $C - 1$ autovectores que corresponden a los autovalores distintos de cero. Usar estos autovectores para formar la matriz de proyección $W = [\omega_K \dots \omega_{C-1}]$, que será usada para obtener los vectores característicos $\Omega_i = W^T x_m^i, m = 1, \dots, N; i = 1, \dots, C$ y $\Omega_{test} = W^T X_{test}$

Teniendo el caso anteriormente descrito: dado el conjunto de entrenamiento compuesto por C clases, donde la i -ésima clase contiene N_i muestras, y dado x_m^i un vector columna d -dimensional, que denota la m -ésima muestra de la clase i -ésima. Si todas las muestras en cada clase son proyectadas sobre el espacio nulo de la matriz de dispersión dentro de la clase S_W (aquel que generan los autovectores asociados a autovalores nulos de la matriz S_W), $N(S_W)$, dan lugar a un único vector denominado vector común

$$x_{\text{com}}^i = P^{(2)} x_m^i, \quad i=1, \dots, C, \quad m=1, \dots, N_i$$

donde $P^{(2)}$ es el operador de proyección ortonormal de $N(S_W)$, surge una pregunta natural; si los vectores comunes x_{com}^i , $i = 1, \dots, C$, son distintos, es decir, si cada uno de esos vectores pueden ser asociados únicamente a la clase i -ésima – o, dicho de otra manera, si hay una correspondencia uno a uno entre los vectores comunes y las clases. Por si este no fuera el caso, por ejemplo, si $x_{\text{com}}^i = x_{\text{com}}^j$, para algún $i \neq j$, entonces el método DCV no podrá ser discriminativo entre las dos clases i y j , lo cual hace este método menos útil.

Planteamos la siguiente condición necesaria para que los vectores comunes sean distintos. Dado $i \neq j$; para que los vectores comunes x_{com}^i , x_{com}^j sean distintos es necesario que las muestras x_m^i , x_n^j en las correspondientes dos clases, sean tales que no puedan encontrar números reales α_m , β_n que satisfacen $\sum_{m=1}^{N_i} \alpha_m = 1$, $\sum_{n=1}^{N_j} \beta_n = 1$, y tal que

$$\sum_{m=1}^{N_i} \alpha_m x_m^i = \sum_{n=1}^{N_j} \beta_n x_n^j, \quad (20)$$

Desgraciadamente, esta condición no constituye una condición suficiente para que los vectores comunes sean distintos, de manera que es necesario imponer una condición de separación lineal de las clases consideradas para que si lo sea. Finalmente, podemos afirmar que suponiendo las clases $i=1, \dots, C$ linealmente separables; entonces los correspondientes vectores comunes son distintos. De esta afirmación se deriva que si todas las muestras x_m^i , $i=1, \dots, C$, $m=1, \dots, N_i$, son linealmente independientes, entonces los vectores comunes x_{com}^i , $i=1, \dots, C$, son distintos.

Si los vectores comunes son distintos, entonces también lo son claramente los vectores comunes discriminativos. Las condiciones anteriores son típicamente satisfechas para los conjuntos de datos en espacios de muestra de alta dimensión. Por ejemplo, para un problema típico de reconocimiento facial con imágenes en escala de grises con 256 niveles de tamaño $128 * 128$, el volumen del espacio de muestra es $(16384)^{256}$. Como la dimensión es muy elevada, es muy probable que las muestras del conjunto de entrenamiento sean linealmente independientes, y por tanto el método DCV puede aplicarse con seguridad para el reconocimiento de muestras.

2.3. Vectores Comunes Discriminativos basados en núcleo (*Kernel DCV*)

En este método, primero se proyecta no linealmente el espacio de entrada original a un espacio característico implícito de mayor dimensión, en el cual se espera que los datos sean linealmente separables. Entonces los vectores de proyección óptima se calculan en este espacio transformado. Ha sido demostrado en [52] que la comparación entre métodos basados en kernels resulta, en general, favorable al método kernel elegido para evaluación.

En algunos casos, puede que los métodos lineales no proporcionen un poder discriminante suficiente para clasificar linealmente clases no separables (por ejemplo, el problema XOR). Es por esto, por lo que recientemente se han propuesto técnicas de análisis de discriminante que utilizan kernels [53]-[55]. Su principal idea es transformar los datos de entrada en un espacio de dimensión mayor mediante una función de proyección no lineal y después aplicar las técnicas de análisis de discriminante lineal en ese espacio. Estos métodos se formulan en términos de productos escalares de las muestras proyectadas, y las funciones kernel se utilizan para calcular estos productos punto. Por lo tanto, la función de proyección no lineal y las muestras proyectadas no se utilizan explícitamente, lo que hace que sean métodos computacionalmente factibles. Sin embargo, el problema de la singularidad de la que participan las matrices normalmente se encuentra en estas aproximaciones ya que la dimensión del espacio proyectado es habitualmente mayor que el tamaño del conjunto de entrenamiento (en particular, este problema siempre se plantea si se utilizan los kernels de Gauss). Se han adoptado dos técnicas diferentes para solventar este problema. Mika y colaboradores usan el criterio original FLDA en el espacio no linealmente proyectado y suman una pequeña matriz de dispersión a la complicada matriz singular [53]. Yang y colaboradores usan el criterio FLDA modificado en el espacio proyectado [55]. Primero proyectan los datos sobre el rango de la matriz de dispersión total de las muestras

proyectadas usando el método kernel PCA [56]; después aplican el método LDA, que maximiza el criterio FLDA modificado en este espacio reducido. La primera de estas aproximaciones se denomina método kernel de análisis de discriminante de Fisher (kernel FDA), y la última aproximación se denomina método kernel PCA + LDA (KPCA+LDA).

Tras estos métodos, un nuevo método fue propuesto [52], el método **kernel DCV**, que aplica el método DCV, descrito anteriormente, en el espacio de dimensión mayor ζ , al que se llega mediante una transformación no lineal. Las características óptimas para discriminación son extraídas del espacio de dimensión mayor no linealmente transformado, ya que se garantiza el criterio FLDA modificado para alcanzar el máximo valor cuando se usa el método kernel DCV, así como el método DCV.

En algunos casos, la dimensión del espacio de muestra puede que no sea lo suficientemente grande para asegurar que los vectores comunes discriminativos sean distintos. Hay tres aproximaciones básicas para tratar con esta situación. Primero, podemos descartar todas las muestras dependientes. Una segunda opción consiste en sumar nuevos vectores de proyección de fuera del subespacio discriminante óptimo a los vectores de proyección que generan el subespacio discriminante óptimo. En este caso, como los nuevos vectores de proyección serán de $R(S_W)$, los vectores característicos ya no producirán los mismos vectores comunes discriminativos. Por lo tanto, el 100% de la precisión en el reconocimiento ya no está garantizada, ya que algunas muestras de entrenamiento podrían resultar inclasificadas en este caso. Una tercera solución sería proyectar las muestras de entrenamiento dentro de un espacio de mayor dimensión, como en el método kernel DCV, método que vamos a evaluar.

Cuando aplicamos el método kernel DCV, primero proyectamos el espacio de muestra original hacia un espacio de dimensión mayor ζ , donde los nuevos vectores comunes discriminativos en el espacio proyectado son distintos entre sí. Esto es debido a que una transformada $\Phi: R^d \rightarrow \zeta$ puede proyectar dos vectores que son linealmente dependientes en el espacio de muestra original en dos vectores que son linealmente independientes en ζ . Notar que el espacio proyectado podría tener arbitrariamente dimensión grande, posiblemente infinita, lo que sugiere el uso del método DCV. Tsuda [57] también demostró que si la matriz kernel K , dada en (22) es definida positiva, entonces todas las muestras proyectadas son linealmente independientes. Por lo tanto,

aunque las muestras de datos puedan ser linealmente dependientes en el espacio de muestra original, se garantiza que los vectores comunes discriminativos sean distintos en ζ siempre y cuando la matriz kernel K sea definida positiva. Por lo tanto, se puede obtener una tasa de reconocimiento del 100% para clases no separables linealmente cuando se aplica el método lineal DCV en ζ .

En el espacio transformado, la matriz de dispersión dentro de la clase, S_W^Φ , es típicamente singular debido a la alta dimensión del espacio proyectado. Así los vectores de proyección óptima que maximizan el espacio nulo basado en el criterio FLDA están en la intersección del espacio nulo $N(S_W^\Phi)$ de S_W^Φ y el rango $R(S_T^\Phi)$ de S_T^Φ . Similar al caso lineal, existen principalmente dos aproximaciones para calcular estos vectores de proyección. Podemos primero proyectar las muestras del conjunto de entrenamiento sobre $N(S_W^\Phi)$ y después aplicar PCA; o podemos primero aplicar PCA para proyectar las muestras del conjunto de entrenamiento sobre $R(S_T^\Phi)$ y después encontrar una base ortonormal para el nuevo espacio nulo de la matriz de dispersión dentro de la clase de las muestras transformadas. Sin embargo, la primera aproximación no es posible ya que los algoritmos que siguen esta aproximación usan la función proyección Φ explícitamente. Por tanto, el segundo enfoque es más apropiado. Las muestras del conjunto de entrenamiento pueden ser proyectadas sobre $R(S_T^\Phi)$ a través del kernel PCA. Después podemos encontrar los vectores que generan el espacio nulo de la matriz de dispersión dentro de la clase de las muestras transformadas. Consecuentemente, obtenemos los vectores comunes discriminativos que representan cada clase. A continuación, resumimos el algoritmo.

ALGORITMO

- **Paso 1**

Proyectar las muestras del conjunto de entrenamiento sobre $R(S_T^\Phi)$ a través de kernel PCA. Dado

$$\tilde{K} = K - 1_M K - K 1_M + 1_M K 1_M = U \Delta U^T \in \Re^{M \times M} \quad (21)$$

Donde Δ es una matriz diagonal de autovalores distintos de cero y U es la matriz de los autovectores normalizados asociados a Δ . Aquí la matriz kernel $K \in \mathbb{R}^{M \times M}$ es dada por $K = \Phi^T \Phi = (K^{ij})_{i,j=1,\dots,C}$, donde cada matriz $K^{ij} \in \mathbb{R}^{N_i \times N_j}$ es definida como

$$K^{ij} = (k_{mn}^{ij})_{\substack{m=1,\dots,N_i \\ n=1,\dots,N_j}} = \langle \phi(x_m^i), \phi(x_n^j) \rangle = k(x_m^i, x_n^j)_{\substack{m=1,\dots,N_i \\ n=1,\dots,N_j}}, \quad (22)$$

Donde $k(\cdot)$ representa la función kernel. La matriz que transforma las muestras del conjunto de entrenamiento sobre $R(S_T^\Phi)$ es $(\Phi - \Phi 1_M)U\Delta^{-1/2}$. Entonces las nuevas matrices de dispersión total y dentro de la clase en el espacio reducido pueden mostrarse como

$$\begin{aligned} \tilde{S}_T^\Phi &= ((\Phi - \Phi 1_M)U\Delta^{-1/2})^T S_T^\Phi (\Phi - \Phi 1_M)U\Delta^{-1/2} = \Delta^{-1/2} U^T U \Delta U^T U \Delta U^T U \Delta^{-1/2} = \Delta \\ \tilde{S}_W^\Phi &= ((\Phi - \Phi 1_M)U\Delta^{-1/2})^T S_W^\Phi (\Phi - \Phi 1_M)U\Delta^{-1/2} = \Delta^{-1/2} U^T \tilde{K}_W \tilde{K}_W^T U \Delta^{-1/2} = \Delta \end{aligned}$$

donde $\tilde{K}_W = K - KG - I_M K + I_M K G = (K - I_M K)(I - G)$.

Siendo $I_M \in \mathbb{R}^{M \times M}$ una matriz con todas las entradas $1/M$; y $G = \text{diag}[G_1, \dots, G_C] \in \mathbb{R}^{M \times M}$ es una matriz diagonal bloque y cada $G_i \in \mathbb{R}^{N_i \times N_i}$ es una matriz con todos sus elementos igual a $1/N_i$.

• Paso 2

Encontrar los vectores que generan el espacio nulo de \tilde{S}_W^Φ . Esto puede ser hecho mediante descomposición. Los autovalores normalizados correspondientes a autovalores nulos de \tilde{S}_W^Φ forman una base ortonormal para el espacio nulo de \tilde{S}_W^Φ . Dado V la matriz cuyas columnas son los autovectores calculados, correspondientes a los autovalores nulos tal que,

$$V^T \tilde{S}_W^\Phi V = 0. \quad (25)$$

- **Paso 3 (Opcional)**

Eliminar el espacio nulo de $V^T \tilde{S}_B^\Phi V$ si existe y rotar las direcciones de proyección de manera que las nuevas matrices de dispersión total y entre clases sean diagonales (es decir, las matrices de dispersión de los vectores característicos de las muestras del conjunto de entrenamiento están incorreladas). Esto es

$$V^T \tilde{S}_B^\Phi V = V^T \tilde{S}_T^\Phi V = V^T \Delta V = L \tilde{\Delta} L^T. \quad (26)$$

La matriz de proyección W será entonces

$$(\Phi - \Phi 1_M) U \Delta^{-1/2} V L. \quad (27)$$

Hay como mucho $C - 1$ vectores de proyección. El vector característico Ω_{test} de una muestra de test es obtenido como

$$\Omega_{test} = W^T (\Phi(x_{test}) - \mu^\Phi) \quad (28)$$

donde W es la matriz cuyas columnas son los vectores de proyección $w_j, j = 1, \dots, C - 1$. Entonces cada entrada del vector característico de una muestra de test puede ser obtenido mediante

$$\begin{aligned} \langle w_j, \phi(x_{test}) - \mu^\Phi \rangle &= \langle w_j, \phi(x_{test}) - \Phi 1_M^T \rangle \\ &= (U \Delta^{-1/2} V L)^T (K_{test} - K 1_M^T - 1_M K_{test} + 1_M K 1_M^T) \end{aligned} \quad (29)$$

donde $1_M^T \in \mathbb{R}^{M \times 1}$ es un vector con todos los términos iguales a $1/M$ y $K_{test} \in \mathbb{R}^{M \times 1}$ es un vector con entradas $\langle \Phi(x_m^i) \Phi(x_{test}) \rangle_{i=1, \dots, C; m=1, \dots, N_i}$

Todas las propiedades matemáticas del DCV lineal son trasladadas al método kernel DCV con las modificaciones que ahora se aplican a las muestras proyectadas, $\Phi(x_m^i), i = 1, \dots, C, m = 1, \dots, N_i$, en ζ . Después de realizar la extracción de características, todas las muestras del conjunto de entrenamiento en cada clase suelen dar lugar a distintos vectores comunes discriminativos. Por lo tanto, como en el caso lineal DCV, está garantizada una precisión de reconocimiento del 100% con respecto a los datos de entrenamiento para este método. En la práctica, si no podemos encontrar fácilmente las funciones kernel que garantizan que los vectores comunes

discriminativos en ζ sean distintos, podemos añadir nuevos vectores del exterior del subespacio discriminante óptimo, sin embargo, en nuestra experiencia, fue muy raro que alguno de los kernels expusiera este problema y, en particular, nunca se observó que los kernels de Gauss tuvieran este problema.

En nuestro estudio, vamos a usar tres funciones kernel diferentes; en primer lugar usaremos la función de kernel polinómico $k(x, y) = (\langle x, y \rangle)^n$, con grados 2 y 3; un kernel gaussiano $k(x, y) = \exp(-|x - y|^2/q)$ para todos los conjuntos de datos.

2.4. Diferencias entre los métodos DCV Lineal y Kernel DCV

La proyección de muestras en un espacio de dimensión mayor vía función de proyección no lineal Φ tiene algunas ventajas sobre el método lineal DVC. Las diferencias entre los dos métodos pueden resumirse como sigue.

- i) El método DCV extrae características lineales del espacio de muestra original, mientras que el método kernel DCV extrae características de un espacio implícito de mayor dimensión. Es posible extraer características no lineales usando el método kernel DCV ya que el espacio proyectado está relacionado no linealmente con el espacio de muestra original. Adicionalmente, tenemos la flexibilidad de crear diferentes fronteras de decisión no lineales mediante un simple cambio de las funciones kernel. Sin embargo, estas mejoras son conseguidas a expensas de cálculos más intensos.
- ii) El método DCV puede ser aplicado solamente en el caso de pequeño tamaño de muestra, y la dimensión del espacio nulo de la matriz de dispersión dentro de la clase debe ser grande en comparación con el conjunto de entrenamiento para tasas de reconocimiento buenas. Sin embargo, estas limitaciones no son aplicables al método kernel DCV propuesto. Podemos aplicar el método kernel DCV a los conjuntos de datos, en los que el número de muestras es mayor que la dimensión del espacio de muestra, usando funciones kernel que aseguren alta dimensión del espacio proyectado.

Capítulo:

3. Bases de datos públicas para reconocimiento facial

En esta sección se presenta una pequeña investigación sobre las bases de datos utilizadas para reconocimiento facial. Realizaremos un estudio comparativo de las bases de datos de caras humanas disponibles [41], para la evaluación de los algoritmos de reconocimiento facial que se han desarrollado hasta el momento. Se presentarán las distintas bases de datos públicas y a continuación se comentarán sus características.

Se busca comparar las distintas bases de datos, partiendo de las imágenes que tenemos disponibles. Para ello debemos investigar sobre los conjuntos de imágenes de sujetos, a lo que llamamos bases de datos, en un entorno de cámaras controlado y equilibrado. Normalmente estas imágenes son adquiridas en condiciones de laboratorio, y pueden presentar ligeras variaciones en cuanto a iluminación, localización y gestos del sujeto.

A continuación, se realiza una descripción con profundidad de los subconjuntos de las cinco bases de datos utilizadas en este proyecto para la evaluación del comportamiento de los métodos DCV y kernel DCV descritos anteriormente y

presentamos una tabla con las principales características y rasgos que tienen nuestras bases de datos para poder probar la robustez de los diferentes métodos frente a oclusiones, diferentes condiciones lumínicas, reescalado de las imágenes y variedad de poses así como el estudio del comportamiento y análisis comparativo de los algoritmos frente a la variabilidad del número de individuos y del número de muestras utilizadas para entrenar los clasificadores.

Finalmente expondremos algunas de las consideraciones de las bases de datos, es decir, de las muestras de las que disponemos, a tener en cuenta en el reconocimiento facial.

Las cinco bases de datos utilizadas son:

- YALE Face database. Usamos en primer lugar esta base de datos, durante la fase de programación del algoritmo DCV. Ejecutamos el mismo utilizando como muestras las imágenes de la base de datos YALE para comprobar el correcto funcionamiento. A través de los resultados mostrados en [11], podemos asegurarnos que nuestro método ofrece el comportamiento deseado.
- AR Face database. Al igual que la base de datos YALE, la usamos para comprobar el comportamiento y los resultados ofrecidos por el método kernel DCV. Además la usamos para evaluar ambos métodos más adelante. En concreto, a través de esta base de datos evaluamos la robustez frente a oclusiones aprovechando que la base de datos de AR contiene numerosas muestras con elementos como gafas, pañuelos, bufandas...
- ORL-ATT. Con esta base de datos nos cercioramos de que hemos implementado el método kernel DCV correctamente. Es decir, programamos el método y comprobamos que los resultados que nos ofrece cuando lo ejecutamos con muestras de la base de datos AT&T and Olivetti, son similares a los publicados en [12]. De esta manera nos aseguramos que la codificación del método kernel DCV es correcta.
- Georgia Tech Database. Esta base de datos contiene imágenes a color de 50 individuos, con 15 imágenes por cada uno de ellos. En esta base de datos no

encontramos variación entre las imágenes en términos de iluminación ni de oclusiones; de manera que nos va a servir para evaluar el impacto del ruido en el reconocimiento. Vamos a evaluar la robustez de los métodos respecto a ruido Gaussiano y ruido Sal y Pimienta cuando los aplicamos en este conjunto de imágenes.

- IMM-Frontal Face Database. Esta base de datos nos va a servir para evaluar el comportamiento de los diferentes métodos cuando disminuimos la resolución de las muestras. Nos beneficiamos del hecho de que las imágenes en esta base de datos tengan una resolución muy alta. Así, podremos hacer un análisis comparativo del comportamiento de nuestros métodos cuando variamos en una cantidad importante el número de píxeles que representan las imágenes.

Con el fin de construir / entrenar algoritmos de reconocimiento facial que ofrezcan evaluaciones fiables, se necesitan bases de datos de imágenes de caras de tamaño considerable. Muchas bases de datos de cara usadas para fines no comerciales están disponibles en Internet, ya sea gratuitamente o por pequeños honorarios.

Estas bases de datos se registran en diversas condiciones y con diversas aplicaciones en mente. En las secciones siguientes se describen brevemente algunos de las bases de datos disponibles que son ampliamente conocidas y utilizadas.

Con el desarrollo de la visión por ordenador y los modelos de reconocimiento, especialmente con la profunda investigación en detección y animación de caras; muchos centros de investigación han construido diversas bases de datos de caras para realizar experimentos sobre estudios correlativos y probar algoritmos.

Las bases de datos que existen actualmente son casi todas imágenes 2D o vídeos, aplicadas en estudios de detección facial, reconocimiento facial, selección de características faciales, expresión facial e iluminación. Las caras en estas bases de datos varían en raza, edad, sexo, postura, expresión e iluminación. A continuación analizamos algunas de las bases de datos de caras más conocidas, que serán las usadas para la evaluación de los algoritmos de reconocimiento facial DVC y Kernel DVC.

3.1. Base de datos de caras YALE

Existen dos bases de datos de caras con esta denominación [58]; inicialmente se creó la base de datos “Yale Face Database”. Esta base contiene 165 imágenes en escala de grises y formato GIF de un total de 15 individuos. Su tamaño es de 6.4 MB, con imágenes de 320 *253 píxeles. Está compuesta por 11 imágenes por sujeto, cada una con una expresión facial o configuración diferente: luz centrada, con gafas, expresión de felicidad, luz en la parte derecha, sin gafas, con expresión neutra, luz en la parte izquierda, expresión triste, somnoliento, sorprendido y guiñando un ojo. A continuación, en la Figura 3.1 se muestra un ejemplo de las once imágenes tomadas a un individuo:

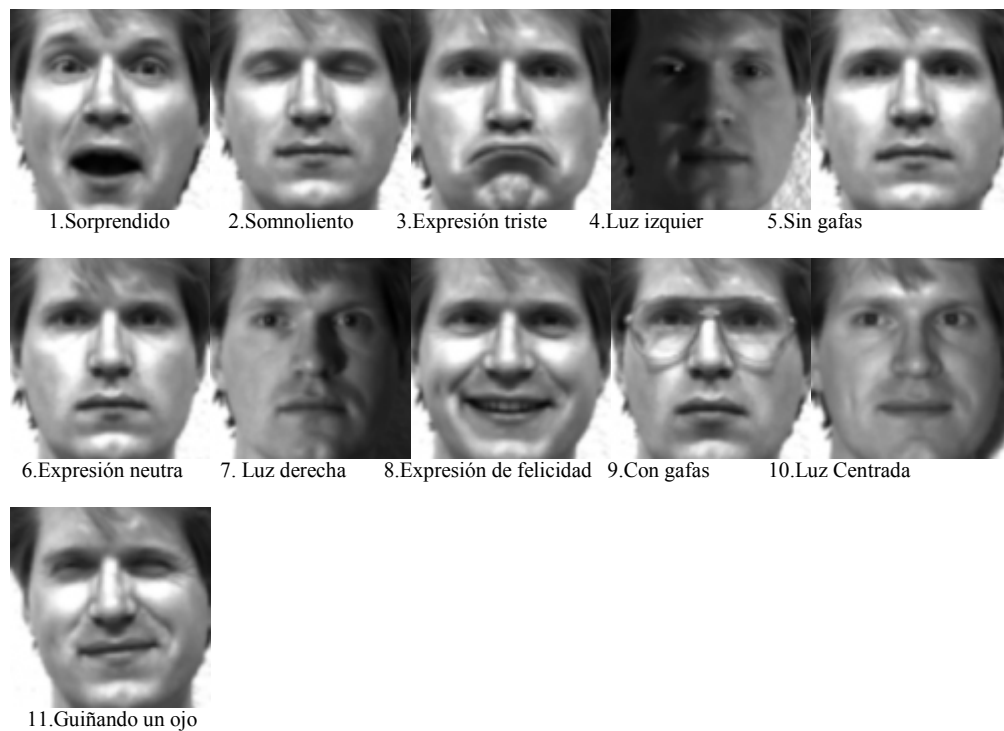


Figura 3.1: Imágenes de una base de datos YALE.

La segunda base de datos es la denominada “Yales Face database B”. Contiene 5760 imágenes, con una única fuente de luz. Contiene imágenes de 10 individuos, cada uno de ellos visto bajo 576 condiciones de visión. (9 posturas x 64 condiciones de iluminación). Para cada sujeto también se toma, en una postura particular, una imagen con luz ambiente. De manera que en total, el número de imágenes asciende a $5760 + 90 = 5850$. El tamaño total de la base de datos es de aproximadamente 1GB. En la Figura 3.2 podemos ver una muestra de varios individuos de esta base de datos.



Figura 3.2: Algunas imágenes de la base de datos YALE B

Las 65 imágenes (64 con diferente iluminación y con luz ambiente) de un individuo en una postura particular han sido comprimidas y almacenadas en un único fichero. Además, disponemos de las coordenadas del centro de cada cara, X e Y y para todos los conjuntos de pose frontal, disponemos de las coordenadas del ojo izquierdo, del ojo derecho y de la boca.

El nombre de fichero está relacionado con el azimut y la elevación de la fuente de luz. Por ejemplo, la imagen con el nombre `yaleB03_P06A+35E+40.pgm` corresponde al sujeto #3 visto desde la pose #6, con una fuente de luz cuya dirección se encuentra a 35 grados de azimut respecto del eje de la cámara y con 40 grados de elevación. Un azimut positivo implica que la fuente de luz se encuentra a la derecha del sujeto y un azimut negativo significa que la luz está a la izquierda del individuo. Elevaciones positivas implican sobre el horizonte, mientras que negativas implican bajo el horizonte.

A continuación, en la figura 3.3 podemos ver un ejemplo de un sujeto bajo condición de iluminación frontal y las nueve diferentes poses:

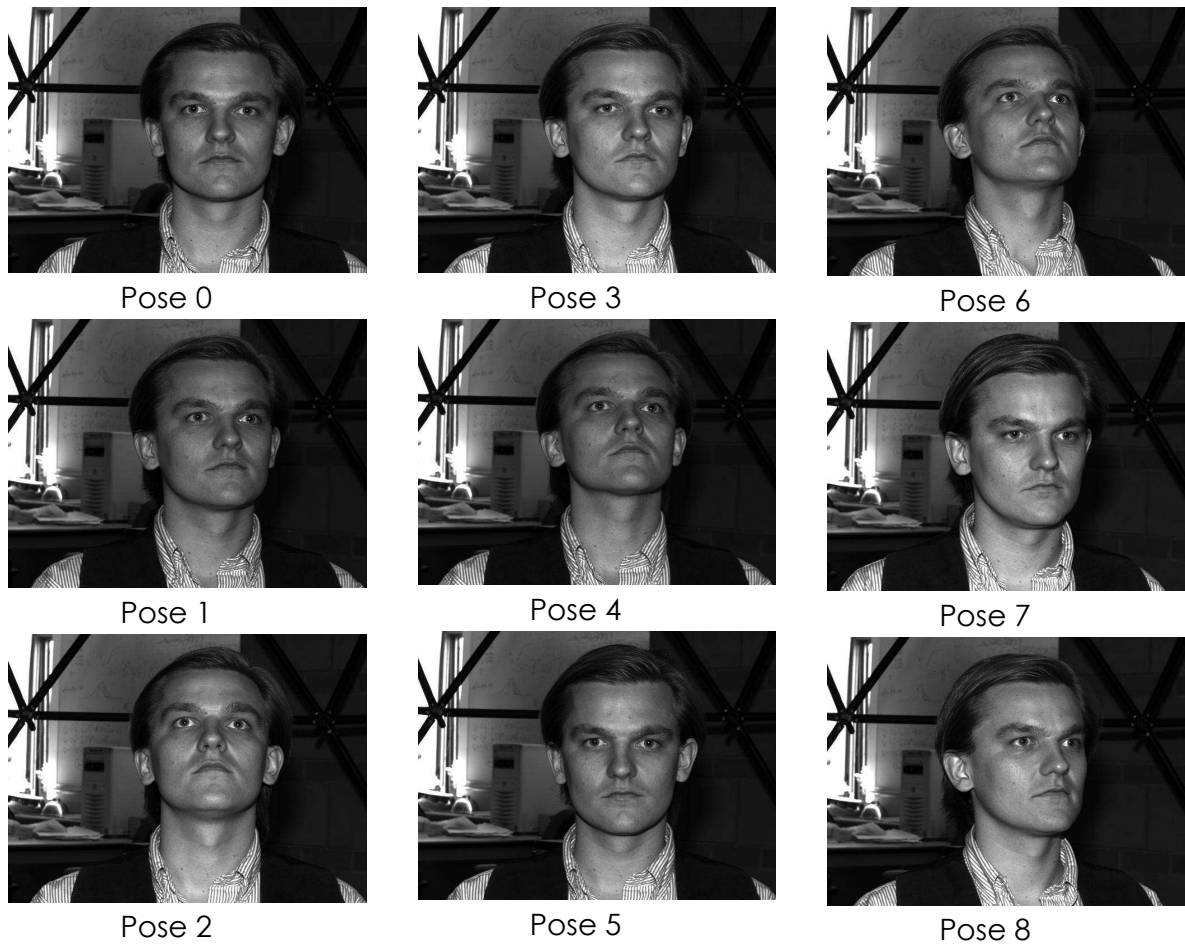


Figura 3.3: Las 9 diferentes poses de un sujeto en la base de datos YALE B.

Las imágenes adquiridas son de 8-bits (escala de grises) y fueron tomadas con una cámara Sony XC-75 (con función de respuesta lineal) y almacenadas en formato PGM. El tamaño de cada imagen es de $640 * 480$.

Finalmente, se decidió evaluar nuestros métodos en la base de datos Yale, descartando el uso de la base de datos Yale B. Decidimos no usarla ya que su adquisición fue bastante dificultosa y las necesidades de almacenamiento eran elevadas. Además, al contener un número no elevado de sujetos, 10, no consideramos relevante su uso en nuestros experimentos.

3.2. Bases de datos de caras AR

Esta base de datos de caras fue creada por Aleix Martínez y Robert Benavente en el Centro de Visión por Computador (CVC) de la UAB en 1998 [59]. Contiene más de 4000 imágenes en color que corresponden a caras de 126 personas (70 hombres y 56 mujeres). Las imágenes ofrecen caras de vista frontal con diferentes expresiones faciales, diferentes condiciones de iluminación, y oclusiones (gafas de sol y pañuelo). Las fotos se tomaron en el CVC bajo condiciones estrictamente controladas. No se les impuso a los participantes ninguna restricción sobre la vestimenta (ropa, gafas...), maquillaje, peinado, etc. Cada persona participó en dos sesiones, separadas en un espacio de tiempo de dos semanas (14 días). Se tomaron las mismas imágenes en ambas sesiones.

Esta base de datos de caras está públicamente disponible y se puede obtener fácilmente de la página web del centro. Las imágenes están almacenadas en 10 CD-ROMs como ficheros RGB RAW (información de pixel). Las imágenes son de 768 x 576 píxeles y de 24 bits de profundidad. Se grabaron las imágenes de hombres como “M-xx-yy.raw” y las de mujeres como “F-xx-yy.raw”; donde ‘xx’ es el identificador del individuo (de ‘00’ a ‘70’ para hombres y de ‘00’ a ‘56’ para mujeres) e ‘yy’ especifica las características de cada imagen, su significado se muestra en la siguiente Figura:

1	Expresión neutral
2	Sonrisa
3	Enfado
4	Grito
5	Iluminación izquierda
6	Iluminación derecha
7	Iluminación total
8	Con gafas de sol
9	Con gafas de sol e iluminación izquierda
10	Con gafas de sol e iluminación derecha
11	Con pañuelo
12	Con pañuelo e iluminación izquierda
13	Con pañuelo e iluminación derecha
14-26	Segunda sesión (con las mismas condiciones que 1-13)

Tabla 3.4: Relación del número de imagen con las características de la mismas en la base de datos AR

A continuación, podemos ver un ejemplo de las imágenes tomadas de un individuo en una sesión con las diferentes características anteriormente numeradas y en el mismo orden.



Figura 3.5: Ejemplo de una sesión de imágenes para un individuo en la base AR Face DAtabase.

3.3. Base de datos de caras AT&T AND OLIVETTI

Esta base de datos es una de las más sencillas y contiene 400 imágenes listas para usar, tomadas entre Abril de 1992 y Abril de 1994 en los Laboratorios AT&T [60]. Está confeccionada a partir de 40 sujetos, de los cuales se tienen 10 imágenes por cada uno de ellos. Estas imágenes están tomadas en escala de grises, de un tamaño de 92 * 112 píxeles y corresponden a retratos que varían en el tiempo, en la iluminación, en la expresión (ojos abiertos, cerrados, sonriendo, serios...) y algunas incluso con gafas o barba; además con poses frontales y con ligera inclinación de cabeza. Todas ellas tomadas con un fondo oscuro y homogéneo donde los sujetos permanecen erguidos. A continuación, podemos observar, en la Figura 3.6, una imagen resumen de las caras en esta base de datos.

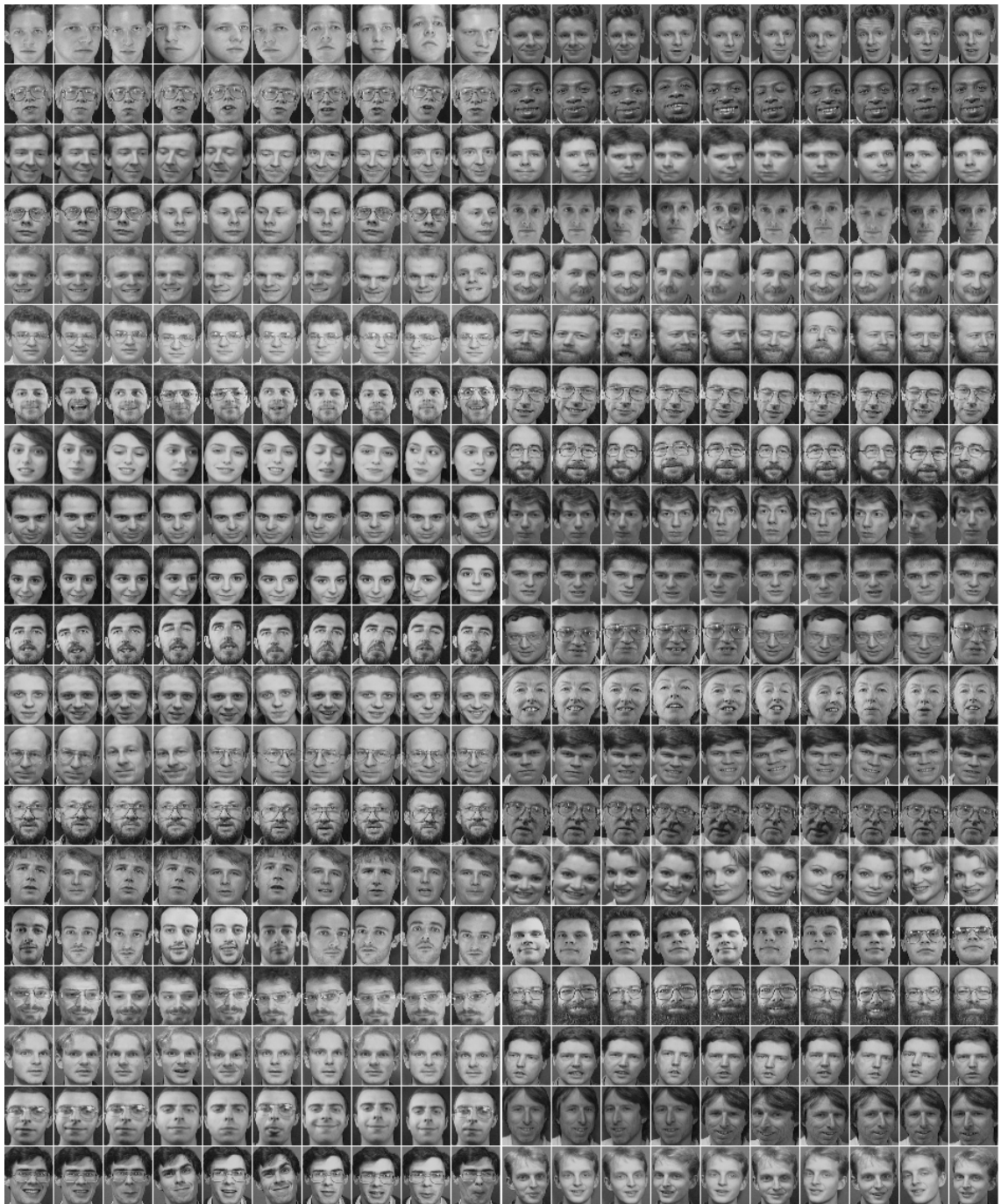


Figura 3.6: AT&T and Olivetti Face Database

3.4. Base de datos de caras GEORGIA TECH

Esta base de datos (128MB) [61] contiene imágenes a color de 50 sujetos tomadas en dos o tres sesiones diferentes en un espacio de tiempo de aproximadamente 5 meses. De cada persona tenemos 15 imágenes con fondo diverso y confuso. La resolución es de $640 * 480$ pixeles. La media del tamaño de los rostros es de $150 * 150$ pixeles. Las imágenes muestran caras de frente y/o inclinadas con diferentes expresiones faciales, condiciones lumínicas y escala. La base incluye hombres y mujeres con variedad de accesorios como pueden ser gafas, gorras, barba,... Cada imagen está etiquetada manualmente para determinar la posición de la cara en la imagen. Además, se proporcionan más detalles en otro documento, que no consideramos para nuestro experimento. A continuación, en la figura 3.7 podemos observar una muestra de esta base de datos, que corresponde a las imágenes tomadas de un individuo en concreto.



Figura 3.7: Georgia Tech Database.

3.5. Base de datos de caras IMM FRONTAL

La base de datos IMM Frontal [62] se registró en 2005 en el Departamento de Informática y Modelización Matemática de la Universidad Técnica de Dinamarca. La base de datos contiene imágenes de 12 personas, todos varones. Han sido registradas durante un periodo de sesiones y se compone de 10 imágenes por cada persona, lo que hace un total de 120 imágenes. Las 10 imágenes de cada persona fueron capturadas en diversas expresiones faciales. Todas las imágenes se grabaron en color con una resolución de $2560 * 1920$ píxeles.



Figura 3.8: Imágenes de un individuo en la base de datos IMM Frontal

3.6. Resumen de las principales bases de datos utilizadas

<i>Propiedades</i>	AR	YALE	IMM Frontal	AT&T and Olivetti	Georgia Tech
# De sujetos	126(70h 56m)	15	12	40	50
# De imágenes	4000	165	10	400	750
Estáticas/Video	Estáticas	Estáticas	Estáticas	Estáticas	Estáticas
Gris/Color	Color	Escala de grises-8 bits	Color	Escala de grises	Color
Resolución	768*576	243*320	2560*1920	92*112	640*480
Posición cara	Vista frontal o cuarto perfil	Variación moderada	Variación moderada	Variación moderada	Vista frontal o cuarto perfil
Expresión Facial	4 expresiones faciales ²	6 expresiones faciales ⁵	Cambios moderados	3 expresiones faciales ⁴	Varias
Iluminación	3 condiciones iluminación ³	3 condiciones de iluminación ³	Variación moderada	N/A	Varias
Accesorios (Oclusión)	Gafas de sol pañuelos	Gafas	Gafas	Gafas	Gafas
Datos 3D	N/A	N/A	N/A	N/A	N/A
Disponible	Si	Si	Si	Si	Si

Figura 3.9: Cuadro-resumen de las principales características de las bases de datos de caras empleadas

- (1) Vista frontal, cuarto perfil izquierdo, cuarto perfil derecho, medio perfil izquierdo, medio perfil derecho, perfil izquierdo y perfil derecho.
- (2) Neutral, sonriente, enfadado y gritando
- (3) Luz derecha, luz izquierda y luz centrada

(4) Neutral, sonriente y con ojos cerrados.

(5) Neutral, feliz, triste, somnoliento, sorprendido y guiñando un ojo

3.7. Algunas consideraciones de las bases de datos en reconocimiento facial

Respecto a las bases de datos que usamos para realizar la evaluación y al proceso de reconocimiento de rostro, debemos tener en cuenta algunas consideraciones generales.

- La variación en los rasgos faciales de las diferentes personas, que puede ser muy pequeña.
- El problema de tener un gran espacio de entrada de dimensiones y sólo un número pequeño de muestras.

El alcance de este proyecto se define con el respeto a estas consideraciones.

Variación en la apariencia facial. Una imagen facial está sujeta a varios factores como pueden ser la pose de la cara, la iluminación y la expresión facial, así como la apertura del objetivo, tiempo de exposición y las aberraciones de la cámara. Debido a estos factores puede ocurrir que tengamos grandes variaciones de imágenes de la cara de la misma persona. Por otro lado, pueden ocurrir a veces pequeñas variaciones interpersonales. Aquí, el extremo es el de gemelos idénticos, como puede verse en la Figura 3.10. Se pueden utilizar diferentes limitaciones en el proceso de adquisición de imágenes para filtrar algunos de estos factores, así como métodos de preprocesamiento.

En una situación en la que la variación entre las imágenes obtenidas de la misma persona es mayor que la variación entre las imágenes de dos individuos, se debe adquirir una mayor cantidad de datos de imágenes 2D de personas para hacer computación basada en reconocimiento facial. Este proyecto se ocupa principalmente de imágenes 2D frontales de caras.



Figura 3.10: Pequeñas variaciones interpersonales ilustradas por gemelos idénticos.

Tomada de www.digitalwilly.com

Análisis facial en un espacio de imagen. Al mirar la información fotométrica de una cara, el reconocimiento facial en su mayoría se basa en el análisis de un subespacio, ya que las caras en las imágenes residen en un subconjunto del espacio de la imagen. Esto puede ser ilustrado por una imagen compuesta de $32 * 32$ píxeles. Esta imagen contiene un total de 1024 píxeles, con la posibilidad de mostrar una gran gama de escenografías diferentes. Usando sólo una escala gris de 8 bits por píxel, la imagen puede mostrar un gran número de configuraciones diferentes, exactamente $256^{1024} = 2^{8192}$. Como comparación, la población mundial es de sólo 2^{32} . Es evidente que sólo una pequeña fracción de estas configuraciones de imagen mostrará caras. Como resultado, la mayor parte de la representación en el espacio original de imagen, es muy redundante desde un punto de vista de reconocimiento facial. Por lo tanto, debe ser posible reducir el espacio de la imagen de entrada para obtener un subespacio mucho más pequeño, donde el objetivo del subespacio es eliminar el ruido y la redundancia, preservando la información discriminativa de la cara.

Gran espacio de entrada y pequeño tamaño de muestra. Otro problema asociado al reconocimiento facial es el elevado espacio de entrada de una imagen y el pequeño tamaño de muestra por lo general de un individuo. Una imagen compuesta de $32 * 32$ píxeles reside en un espacio 1024-dimensional, donde el número de imágenes de una persona en concreto normalmente es mucho menor. Puede que un pequeño número de imágenes de una persona concreta no sea suficiente para hacer una aproximación adecuada de la variedad, lo que puede causar un problema. Un ejemplo de este problema se muestra en la Figura 3.11. En la actualidad, hay solución conocida

para este problema, solución que deriva en el método de reconocimiento facial DCV que evaluamos; aparte de la captura de un número suficiente de muestras para una aproximación satisfactoria.

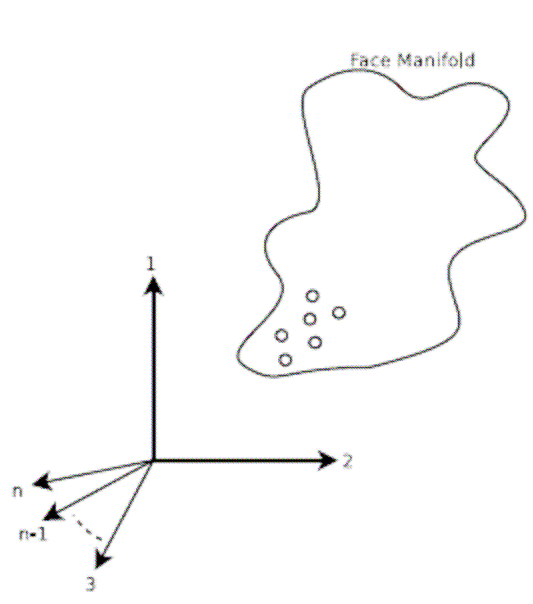


Figura 3.11: Una ilustración del problema de no ser capaz de aproximar satisfactoriamente la variedad cuando solamente tenemos un pequeño número de muestras. Las muestras se denotan mediante círculos.

Capítulo:

4. Evaluación de experimentos

Como inicio del grupo de experimentos realizados, nos encontramos con la tarea de verificación y comprobación de la implementación de los métodos de reconocimiento facial escogidos para nuestro análisis. En primer lugar, tras elaborar el primer método, el método DCV, comparamos los resultados obtenidos con los mostrados en el documento [51]. En este documento, proponen la aplicación del método a dos bases de datos, la base de datos Yale [58] y la base de datos AR [59]. A continuación, abordamos la verificación del segundo método, es decir, comparamos los resultados propios con los obtenidos en el documento [52], que evalúan el método kernel DCV en la base de datos ORL [60].

Tras una primera aproximación a los métodos y su comportamiento en varias bases de datos, procedemos a evaluar la robustez de los métodos objeto de estudio frente a varios de los problemas con los que nos podemos encontrar. Como hemos descrito con anterioridad, estos problemas o dificultades son inherentes a las caras o la forma en que las imágenes han sido capturadas en el mundo real. Evaluaremos, por tanto, los métodos DCV y KDCV en sus tres variantes en torno a los siguientes factores:

- Variación del número de individuos (C) en la base de datos
- Variación del número de muestras de entrenamiento (N_{train})
- Variación en las condiciones de iluminación
- Existencia de oclusiones en las imágenes
- Existencia de ruido (Gaussiano y Sal y Pimienta)
- Variación de la resolución

4.1. Verificación de la implementación de los métodos DCV y KDCV

4.1.1. Verificación de la implementación del método DCV

En [51] ensayan el método DCV en la base de datos Yale ligeramente modificada. Esta base de datos, como se vio en el apartado anterior, se compone de $C=15$ sujetos y $N=15$, número de imágenes por sujeto. Para los sujetos numerados 2, 3, 6, 7, 8, 9 y 14, las imágenes de expresión facial normal y las imágenes sin gafas (o con gafas en el caso de que el sujeto normalmente lleve gafas) fueron copias unas de otras. Por tanto, las imágenes sin gafas (o con gafas si el sujeto normalmente las lleva) de todos los sujetos fueron eliminadas para hacer que todas las clases tengan el mismo número de muestras y todas ellas distintas. Así, se dispone de 10 muestras por sujeto produciendo una base de datos de tamaño $M=150$. Además, en este ensayo se procesan las imágenes alineándolas y escalándolas, de manera que las distancias entre los ojos fueran las mismas para todas las imágenes y se aseguran que los ojos se encuentran en las mismas coordenadas de la imagen. Además la imagen resultante fue recortada. El tamaño de la imagen final fue de $126 * 152$ píxeles. Se calcularon las tasas de reconocimiento mediante la estrategia “dejar uno fuera” (*leave-one-out*) [63] ya que el conjunto de entrenamiento es relativamente pequeño. Para clasificación, se usó el algoritmo “vecino más cercano” (*nearest-neighbor*) utilizando distancia Euclídea.

Nuestro experimento difiere del propuesto ya que no se hace preprocesado de las imágenes, es decir, de aquí en adelante tratamos las imágenes tal y como se presentan en las bases de datos, sin ningún tipo de alineamiento o recorte. La única operación a la que nos vemos obligados a realizar es al reescalado de las imágenes. Esto se hace en

algunas ocasiones en las que las fotografías son muy pesadas y por necesidades computacionales a la hora de utilizar la herramienta con la que realizamos los algoritmos de los métodos, *matlab*; debemos reducir el tamaño. En este caso concreto no tuvimos la necesidad de reescalar, de manera que el tamaño de las imágenes es el original, $320 * 240$ píxeles. Por ello, cabe esperar que los resultados de las tasas obtenidas en nuestros experimentos sean inferiores. Los resultados de la base de datos Yale figuran en la Tabla 4.1.

TABLA 4.1
TASA DE RECONOCIMIENTO CON DCV PAR LA BASE DE DATOS YALE

Método	[51]	PFC
DCV	97.33%	87.33%

Verificamos que esta tasa de reconocimiento para el método lineal es inferior en un 10%. Asumimos que esta diferencia se produce porque nuestras imágenes no han sido previamente procesadas como lo han hecho en el documento.

Además en [51] ensayan con la base de datos AR, que se compone de 126 individuos y 26 imágenes por individuo. Se seleccionan aleatoriamente $C=50$ individuos (30 hombres y 20 mujeres) para el experimento. Solamente se escogieron imágenes sin oclusiones para todos los sujetos (Fig 1a, 1b, 1c, 1e, 1f, y 1g, y 1n, 1o, 1p, 1q, 1r, 1s, 1t). De este modo, el tamaño de la base de datos resultante fue de 700 con 14 imágenes por sujeto. Después, estas imágenes se convirtieron a escala de gris, fueron alineadas, escaladas, centradas y recortadas, utilizando el mismo procedimiento descrito previamente para el experimento para la base de datos Yale. El tamaño final de las imágenes fue de $222 * 229$ píxeles. El conjunto de entrenamiento consistió en $N_{\text{train}}=7$ imágenes de cada sujeto seleccionadas aleatoriamente, y el resto de las imágenes fueron utilizadas para el conjunto de test. Así obtuvimos un conjunto de entrenamiento de $M_{\text{train}}=350$ imágenes. Para clasificación se empleó un algoritmo de “vecino más próximo” utilizando distancia Euclídea. Se repitió este proceso hasta ocho veces y las tasas de reconocimiento fueron halladas como la media de las tasas de reconocimiento de cada ejecución.

En nuestro caso procedemos de la misma manera con el experimento con la salvedad de que no hacemos preprocesado en la imágenes y haciendo una reducción del tamaño hasta tener imágenes de $288 * 384$ píxeles. Como se ha comentado anteriormente, nos vemos obligados a realizar esta disminución de tamaño porque *matlab* presenta serios problemas de memoria cuando tratamos de aplicar la implementación de nuestros métodos. Los resultados con la base AR se encuentran en la Tabla 4.2.

TABLA 4.2
TASA DE RECONOCIMIENTO CON DCV PAR LA BASE DE DATOS AR

Método	[51]	PFC
DCV	99.35%	99.14%

En este caso hemos hallado una tasa mucho más próxima a la original, no influyendo el preprocesado tanto en el reconocimiento como en el experimento con la base de datos Yale. El éxito del método propuesto depende del tamaño del espacio nulo de la matriz de dispersión intra-clase, S_W . Cuando el tamaño del espacio nulo es pequeño, se espera que las tasas de reconocimiento sean malas, ya que no habrá suficiente espacio para obtener los vectores de proyección óptimos. Esto también se menciona en [48].

4.1.2. Verificación de la implementación de los métodos KDCV: con kernel; polinómico de segundo grado, polinómico de tercer grado y gaussiano

Ahora procedemos a comprobar la implementación del segundo método, el método Kernel DCV, con sus diferentes variantes escogidas, el método kernel DCV con kernel polinómico de segundo grado ($n=2$), con kernel polinómico de tercer grado ($n=3$) y con kernel gaussiano; así como el método DCV. Utilizamos como referencia para comprobar los resultados obtenidos la publicación [52].

Una selección apropiada de las funciones kernel para tareas especiales es todavía un problema abierto ya que diferentes funciones de kernel dan lugar a diferentes construcciones de espacio característico implícito [64]. Usamos los siguientes kernels:

- Polinómicos de grado $n=2$: $k(x,y) = (\langle x,y \rangle)^2$
- Polinómico con grado $n=3$: $k(x,y) = (\langle x,y \rangle)^3$
- Gaussiano: $k(x,y) = \exp(-\|x - y\|^2/q)$

para todos los conjuntos de datos.

En el procedimiento que llevan a cabo en la publicación [52], se emplea un pequeño conjunto creado aleatoriamente a partir de los conjuntos de entrenamiento y test para calcular el parámetro Gaussianos, q , que mejores la base de datos. Estos conjuntos de datos se usaron únicamente para la selección del parámetro y no se usaron para probar el comportamiento de generalización de los métodos. Primero se calculan los valores máximo y mínimo de los parámetros gaussianos que producen tasas de reconocimiento aceptables mediante una búsqueda global sobre el amplio rango del espacio de parámetros. Después, se divide linealmente el intervalo determinado por los valores máximo y mínimo de los parámetros en algunos subintervalos y se calculan las tasas de reconocimiento. Finalmente, se lleva a cabo una búsqueda en el vecindario de los parámetros Gaussianos que producen la mejor tasa de reconocimiento y se calcula el

mejor parámetro Gaussiano final. En este documento, se realizan experimentos sobre la base de datos Olivetti-Oracle Research Lab (ORL) [60], que se compone de $C=40$ sujetos con $N=10$ muestras por sujeto. El tamaño de cada imagen es de $320 * 240$ píxeles. Aleatoriamente se seleccionan $N_{\text{train}}=5$ muestras de cada clase para entrenamiento y el resto ($10-N_{\text{train}}$) de las muestras de cada clase se usaron para test. Este proceso se repitió 45 veces, y se crearon 45 conjuntos de test y entrenamiento diferentes. Los primeros cinco conjuntos de datos (de entrenamiento y test) se usaron para la selección de parámetros, el resto se usaron para la evaluación del comportamiento. No se aplicó ningún preprocesado a las imágenes. Se comparó el vector característico de la muestra de test solamente con los vectores comunes discriminativos mediante el uso de distancia Euclídea para estos métodos. Las tasas de reconocimiento del experimento fueron encontradas mediante la media de las tasas de reconocimiento de cada prueba de las 40 realizadas. Los resultados ofrecidos por el estudio y los nuestros se pueden ver en la Tabla 4.3.

TABLA 4.3
TASA DE RECONOCIMIENTO PAR LA BASE DE DATOS ORL

Numero de muestras de entrenamiento en cada clase	Tasas de Reconocimiento (%) y Desviaciones Típicas							
	DCV		Kernel DCV n=2		Kernel DCV n=3		Kernel Gaussiano	
	[52]	PFC	[52]	PFC	[52]	PFC	[52]	PFC
N = 5	96.11	96.22	96.32	96.14	95.45	95.73	96.71	96.81
	$\sigma=1.73$	$\sigma=1.41$	$\sigma=1.68$	$\sigma=1.51$	$\sigma=1.83$	$\sigma=1.53$	$\sigma=1.97$	$\sigma=1.44$

Podemos observar que las tasas conseguidas mediante nuestro experimento son prácticamente iguales a las obtenidas en la publicación [52] mediante el mismo procedimiento. Es importante mencionar que los valores del parámetro q en ambos casos son diferentes. En el documento referenciado este parámetro es $q = 1.06e8$ para todos los casos, mientras que en nuestro experimento hemos hallado un valor de $q = 4.4e9$. El procedimiento para hallar este parámetro ha sido similar al mencionado, acotando un conjunto inicial de búsqueda del parámetro de entre $1/4 \cdot d^2$ y $4 \cdot d^2$, siendo d la dimensión del espacio de muestra, $d = 92 * 112 = 10304$. Conseguimos una tasa de reconocimiento superior cuando utilizamos el KDCV con kernel gaussiano. Aunque

teniendo en cuenta los valores de varianza podemos considerar que las tasas son similares.

4.2. Robustez de los métodos DCV y KDCV frente a la variación del número de individuos en la base de datos

El éxito de los métodos propuestos depende del tamaño del espacio nulo de la matriz de dispersión dentro de la clase S_w . Cuando el tamaño del espacio nulo es pequeño, se espera que las tasas de reconocimiento sean malas, ya que no habrá suficiente espacio para obtener los vectores de proyección óptimos.

Para verificar este hecho, ejecutamos nuestros experimentos usando las imágenes de la base de datos ORL [60], ya que dispone de un número elevado de clases ($C=40$) y el tamaño de sus imágenes no nos supone una limitación ($92 * 112$ píxeles), por lo que no es necesario reescalarlas; además las imágenes son en escala de gris. Aleatoriamente seleccionamos cinco imágenes de cada clase para entrenamiento y usamos las cinco restantes para test. Así, obtuvimos un conjunto de entrenamiento de 200 imágenes y uno de test de 200 imágenes también.

Para observar el descenso del rendimiento debido a un espacio nulo pequeño, idealmente tendríamos que tener un gran número de clases para el conjunto de entrenamiento con un tamaño del espacio de muestra de $92 * 112$ píxeles, pero podemos apreciar con el número de clases de las que disponemos, este hecho. Este procedimiento se repitió hasta ocho veces, y las tasas de reconocimiento se calcularon como la media de las tasas de cada ejecución. Los resultados se muestran en la Tabla 4.4 y su representación la observamos en la Gráfica 4.1. Como se puede observar, el rendimiento decrece con la reducción de la dimensión del espacio nulo.

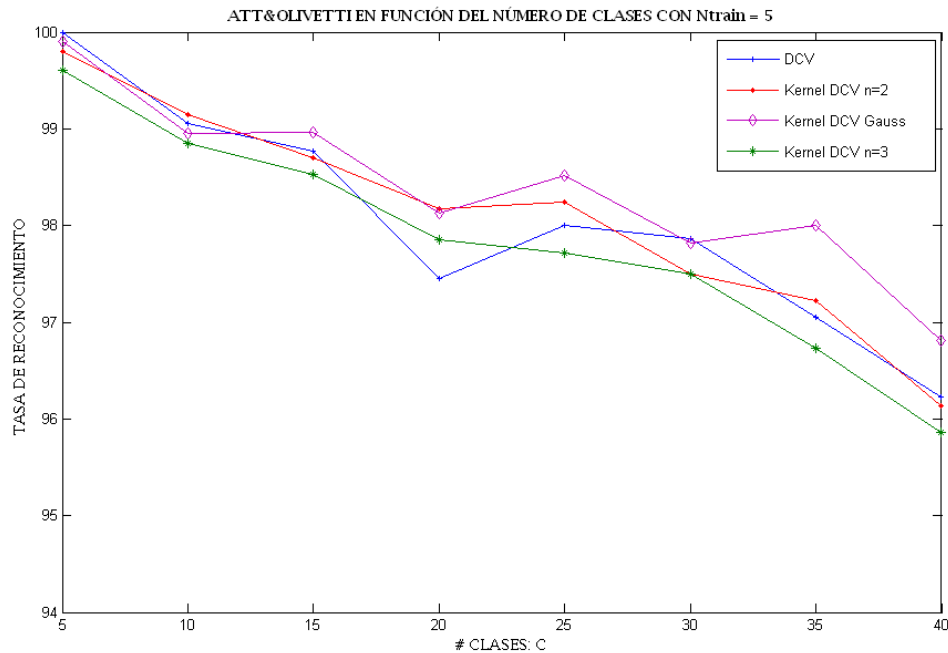
TABLA 4.4

TASA DE RECONOCIMIENTO PAR LA BASE DE DATOS ORL CON DIFERENTE
NÚMERO DE INDIVIDUOS EN LA BASE DE DATOS

Métodos utilizados con $N_{\text{train}}=5$	Tasas de Reconocimiento (%) y Desviaciones Típicas NÚMERO DE INDIVIDUOS							
	C=5	C=10	C=15	C=20	C=25	C=30	C=35	C=40
DCV	100 $\sigma=0$	99.05 $\sigma=1.19$	98.77 $\sigma=1.10$	97.45 $\sigma=1.78$	98.00 $\sigma=1.44$	97.87 $\sigma=1.81$	97.06 $\sigma=1.72$	96.22 $\sigma=1.41$
Kernel DCV n=2	99.80 $\sigma=0.88$	99.15 $\sigma=1.19$	98.70 $\sigma=1.61$	98.17 $\sigma=1.17$	98.24 $\sigma=1.01$	97.50 $\sigma=1.57$	97.23 $\sigma=1.88$	96.13 $\sigma=1.52$
Kernel DCV n=3	99.60 $\sigma=1.21$	98.85 $\sigma=1.49$	98.53 $\sigma=1.53$	97.85 $\sigma=1.29$	97.72 $\sigma=1.24$	97.50 $\sigma=1.51$	96.73 $\sigma=1.27$	95.73 $\sigma=1.53$
Kernel Gaussiano	99.90 $\sigma=0.63$	98.95 $\sigma=1.01$	98.97 $\sigma=1.11$	98.12 $\sigma=1.28$	98.52 $\sigma=1.27$	97.82 $\sigma=1.60$	98.00 $\sigma=1.32$	96.81 $\sigma=1.44$

GRAFICA 4.1

TASA DE RECONOCIMIENTO PAR LA BASE DE DATOS ORL CON DIFERENTE
NÚMERO DE INDIVIDUOS EN LA BASE DE DATOS



En primer lugar, observamos que las mejores tasas se consiguen en casi todos los casos con kernel gaussiano. Cuando tenemos un número pequeño de individuos en la

base de datos obtenemos tasas de entorno al 100%. Es decir, al disponer de pocos individuos, el número de autovectores no nulos en la matriz de dispersión S_W sería también pequeño ($C-1$), por lo que la dimensión de espacio nulo de S_W , $N(S_W)$ sería elevada y esto produciría mejores tasas de reconocimiento. Concretamente, el reconocimiento es total en el caso del método DCV, mientras que en el resto las tasas son muy elevadas pero no alcanzan el 100%.

A medidas que aumentamos el número de individuos en la base de datos, disminuyen las tasas de reconocimiento hasta entorno a un 4% para el caso de mayor número de clases. El mejor comportamiento lo realiza el método KDCV con kernel Gaussiano, siendo mayor la diferencia con respecto a los demás métodos cuando aumentamos C . El resto de los métodos se comportan de manera similar, siendo un poco más elevada la tasa del método KDCV polinómico con $n=2$ que la del método KDCV polinómico con $n=3$.

Los que más decrementan su rendimiento cuando aumentamos el número de individuos en la base de datos son los métodos lineal y con kernel polinómico de $n=3$. El método lineal pasa de proporcionar reconocimiento total (100%) para 5 individuos en la base de datos a tener un 96.22% de aciertos cuando la base de datos es de 40 sujetos. Es decir, ve su tasa reducida en un 3.78%. El método de kernel polinómico de tercer grado lo hace en un 3.87%, pasando a tener una tasa de 95.73% para el caso de mayor número de individuos posible. En menor medida lo hace el método con kernel gaussiano, que lo hace en un 3%.

Los métodos que peores tasas proporcionan cuando la base de datos utilizada contiene 40 individuos son los métodos KDCV con kernels polinómicos. Éstos reducen sus tasas llegando en el peor de los casos a tener un 95.73% el KDCV con $n=3$ y 96.13% el método con $n=2$.

4.3. Robustez de los métodos DCV y KDCV frente a la variación del número de muestras

En este caso también decidimos usar la base de datos ORL [60] ya que tiene un número de muestras relativamente alto, 10 imágenes por individuo y 40 sujetos, y además tenemos como referencia el documento [52], en el que también se hallan tasas de reconocimiento con diferente número de muestras de entrenamiento.

Para ello escogemos hasta 10 conjuntos aleatorios de muestras de entrenamiento con $N_{\text{train}} = 3, 5$ y 7 . Es decir, realizamos el proceso de reconocimiento diez veces con cada número de muestras de entrenamiento citado y hallamos las tasas finales como promedio de las 10 ejecuciones en cada caso. A las imágenes no les realizamos ningún tipo de preprocesado (en el documento con el que comparamos tampoco se hace procesado alguno a las imágenes), tampoco llevamos a cabo reescalado de las imágenes ya que su tamaño, 92×112 píxeles, no nos supone un inconveniente a la hora de utilizar *matlab* para aplicar los métodos.

Las tasas recogidas junto con las publicadas en [52] se muestran a continuación en la Tabla 4.5 y se representan en la Gráfica 4.2

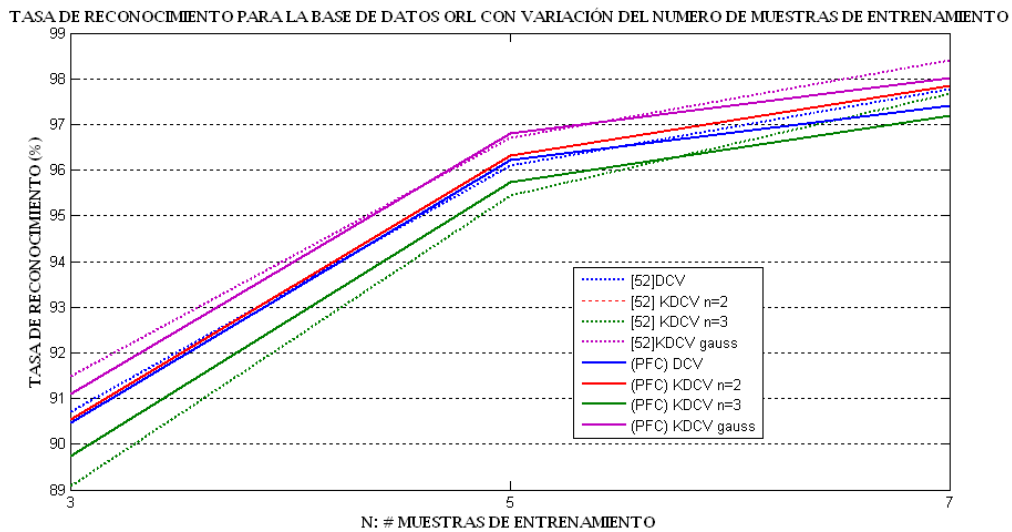
TABLA 4.5

**TASA DE RECONOCIMIENTO PAR LA BASE DE DATOS ORL CON DIFERENTE
NÚMERO DE MUESTRAS DE ENTRENAMIENTO**

Numero de muestras de entrenamiento en cada clase	Tasas de Reconocimiento (%) y Desviaciones Típicas							
	DCV		Kernel DCV n=2		Kernel DCV n=3		Kernel Gaussiano	
	[52]	PFC	[52]	PFC	[52]	PFC	[52]	PFC
N = 3	90.70 $\sigma=2.49$	90.46 $\sigma=2.49$	90.54 $\sigma=2.72$	90.22 $\sigma=2.81$	89.09 $\sigma=2.93$	89.73 $\sigma=1.65$	91.50 $\sigma=2.94$	91.09 $\sigma=2.88$
N = 5	96.11 $\sigma=1.73$	96.22 $\sigma=1.41$	96.32 $\sigma=1.68$	96.14 $\sigma=1.51$	95.45 $\sigma=1.83$	95.73 $\sigma=1.53$	96.71 $\sigma=1.97$	96.81 $\sigma=1.44$
N = 7	97.77 $\sigma=1.33$	97.42 $\sigma=1.43$	97.85 $\sigma=1.46$	97.81 $\sigma=1.19$	97.67 $\sigma=1.53$	97.20 $\sigma=1.65$	98.40 $\sigma=1.18$	98.02 $\sigma=1.21$

GRAFICA 4.2

**TASA DE RECONOCIMIENTO PAR LA BASE DE DATOS ORL CON DIFERENTE
NÚMERO DE MUESTRAS DE ENTRENAMIENTO**



Conseguimos una tasa de reconocimiento superior cuando utilizamos el KDCV con kernel gaussiano. En todos los casos en los que tenemos diferente número de muestras de entrenamiento, y como es lógico pensar, obtenemos en todos los casos mejores tasas cuando entrenamos con un número de muestras mayor. El método KDCV con $n = 2$ tiene un comportamiento mejor respecto al lineal y respecto al KDCV $n = 3$, siendo éste algo inferior al lineal para todos los casos.

Para el método con kernel gaussiano conseguimos pasar de una tasa de reconocimiento del 91.5% cuando entrenamos con 3 muestras al 98% cuando entrenamos con 7 muestras, es decir conseguimos incrementar la tasa en un 6.5%. Seguidamente, para 7 muestras, se encuentra el método KDCV, cuyo incremento es mayor, del 7.59%. Este método pasa de proporcionar una tasa del 90.22% a 97.81%. El método lineal padece un incremento del 6.96%, situándose, para un entrenamiento con 7 muestras, en tercer lugar con una tasa de aciertos del 97.42%. El método más deficiente en todos los casos es el método KDCV con $n=3$, que pasa de tener una tasa de reconocimiento del 89.09%, a una tasa del 97.81%. Se puede observar que las diferencias entre los resultados de los diversos métodos parecen disminuir cuando aumentamos el número de muestras que utilizamos en la etapa de entrenamiento.

Respecto a las tasas obtenidas en [52], podemos observar que son muy similares a las obtenidas en nuestro PFC, llegando, incluso, a no observar apenas diferencias en el caso del método KDVC con kernel polinómico de $n=2$.

Además, decidimos representar las tasas de reconocimiento para los métodos estudiados en función del número de muestras de entrenamiento y del número de individuos (clases) en la base de datos a la vez. Repetimos el proceso citado anteriormente; es decir, escogemos hasta 10 conjuntos aleatorios de muestras de entrenamiento con $N_{\text{train}} = 3, 5$ y 7 . Es decir, realizamos el proceso de reconocimiento diez veces con cada número de muestras de entrenamiento y variando el número de individuos en la base de datos, desde 5 hasta 40. Hallamos las tasas finales como promedio de las 10 ejecuciones en cada caso.

Los resultados obtenidos los podemos observar en la Tabla 4.6. Además, representamos los resultados en tres gráficas; Gráfica 4.3, Gráfica 4.4 y Gráfica 4.5, para los diferentes números de muestras de entrenamiento que usamos y ponemos todos los resultados en conjunto en la Gráfica 4.6 para realizar una mejor comparación.

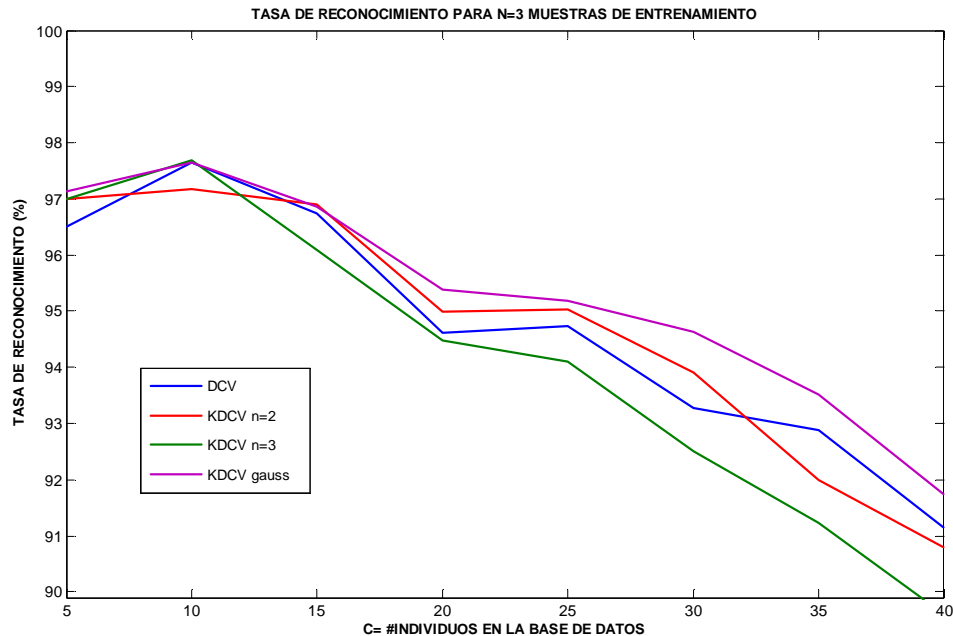
TABLA 4.6

**TASA DE RECONOCIMIENTO PAR LA BASE DE DATOS ORL CON DIFERENTE
NÚMERO DE MUESTRAS DE ENTRENAMIENTO Y DE CLASES**

Número de muestras de entrenamiento en cada clase (N)		Tasas de Reconocimiento (%)							
		Número de individuos en la base de datos (C)							
		C=5	C=10	C=15	C=20	C=25	C=30	C=35	C=40
		Tasa%	Tasa%	Tasa%	tasa%	tasa%	tasa%	tasa%	tasa%
N = 3	DCV	96.50	97.64	96.74	94.61	94.73	93.27	92.88	91.15
	KDCV _{n=2}	97.00	97.18	96.90	94.98	95.03	93.90	92.00	90.80
	KDCV _{n=3}	97.00	97.68	96.09	94.48	94.11	92.50	91.23	89.62
	KDCV _{gauss}	97.14	97.64	96.86	95.39	95.18	94.63	93.52	91.75
N = 5	DCV	99.70	99.00	98.90	97.65	98.30	97.72	97.01	95.81
	KDCV _{n=2}	99.40	99.00	98.97	97.90	98.18	97.85	96.97	96.31
	KDCV _{n=3}	99.60	98.95	99.90	98.22	97.58	97.17	96.88	95.56
	KDCV _{gauss}	99.90	98.60	99.23	98.20	98.06	98.08	97.58	96.40
N = 7	DCV	100	99.25	99.44	98.87	98.93	98.72	98.55	97.75
	KDCV _{n=2}	100	99.67	99.28	98.87	99.37	99.22	98.69	97.46
	KDCV _{n=3}	100	99.33	99.67	98.62	99.17	98.75	98.38	97.62
	KDCV _{gauss}	100	99.17	99.05	98.21	99.07	99.58	99.36	98.48

GRAFICA 4.3

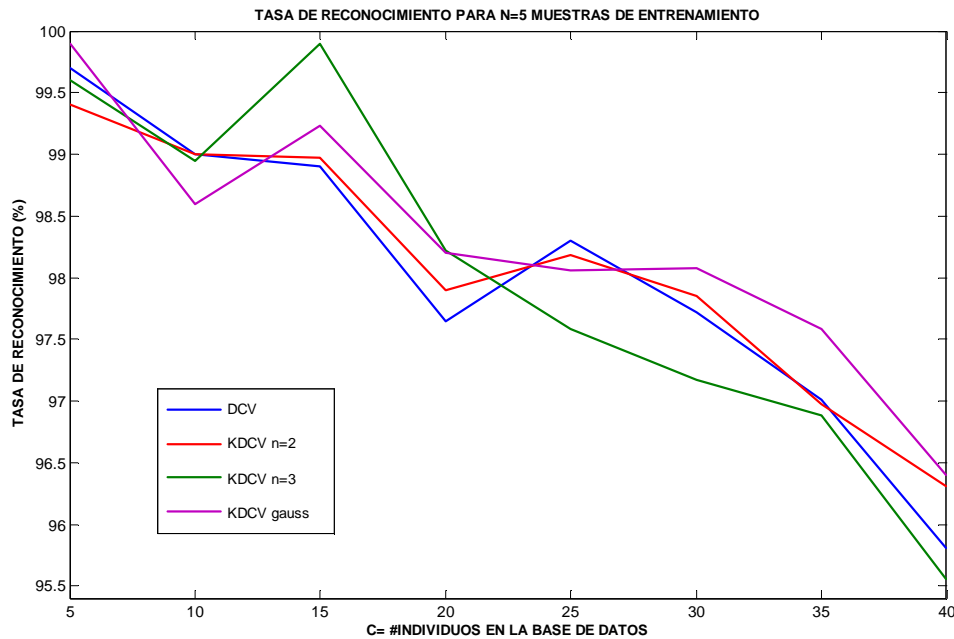
**TASA DE RECONOCIMIENTO PAR LA BASE DE DATOS ORL CON N=3 EN FUNCIÓN
DEL NÚMERO DE INDIVIDUOS**



Quando entrenamos con 3 muestras obtenemos unas tasas de reconocimiento algo bajas. Para todos los casos en los que modificamos el número de clases de la base de

datos, el método que mejor se comporta es el KDCV con kernel gaussiano. Desde que realizamos el experimento con $C=15$ clases hasta que lo hacemos con 40 clases, las tasas de todos los métodos decrecen gradualmente. El método KDCV con kernel gaussiano, que es el que proporciona un rendimiento mayor, ve decrementada su tasa de reconocimiento en este intervalo, pasando de ser 96.86% a 91.75%. Se reduce en un 5.11%. A continuación se sitúa en orden descendiente de rendimiento, el método KDCV con $n = 2$, que se mantiene superior al método lineal mientras en número de clases sea aproximadamente menor que 32. El método lineal es el segundo que mejor tasa de reconocimiento proporciona cuando utilizamos 40 clases, siendo ésta de 91.15%. La diferencia de la tasa de aciertos para este método cuando pasamos de tener 15 clases a 40 clases es de 5.59%. Se reduce algo más que el método con kernel gaussiano, pero aún más lo hacen los métodos KDCV con kernels polinómicos, en un 6.1% el de grado 2 y en un 6.71% el de grado 3. Este último es el que más baja mantiene la tasa de reconocimiento en todo el intervalo, respecto al resto de los métodos; llegando a una tasa de 89.62% cuando tenemos una base datos con 40 sujetos.

Para el intervalo hasta 15 clases, vemos que todos los métodos producen un máximo cuando $C=10$ salvo el método KDCV con $n=3$ que produce una tasa de acierto de 97.2%, el resto de los métodos producen una tasa del 97.7% aproximadamente. El método KDCV con $n=2$ se mantiene constante hasta que evaluamos el método con 15 individuos. Tenemos unas tasas similares, 97%, para todos los métodos cuando empezamos nuestro experimento en una base de datos de 5 individuos a excepción del método DCV, que proporciona una tasa ligeramente inferior del 96.5%.

GRAFICA 4.4**TASA DE RECONOCIMIENTO PAR LA BASE DE DATOS ORL CON N=5 EN FUNCIÓN DEL NÚMERO DE INDIVIDUOS**

Como podemos observar, cuando entrenamos con 5 muestras, el comportamiento es algo más variable. También hay una tendencia de descenso de las tasas de reconocimiento cuando aumentamos el número de clases pero no de una manera tan uniforme. El método KDCV con kernel gaussiano, en este caso, es el que mejores tasas ofrece cuando $C > 20$ aproximadamente, obteniendo tasas muy similares a los métodos KDCV con $n=2$ y DCV en el entorno de $C=25$; aproximadamente del 98%. Anteriormente, para $C < 20$, el método KDCV con $n=3$ es el que mejores tasas produce, llegando a alcanzar un 99.9% para $C=15$. Los métodos DCV y KDCV con $n=2$, en este caso, acusan un comportamiento similar, manteniendo mayoritariamente las menores tasas respecto al resto.

Inicialmente, para 5 individuos se consiguen las mayores tasas con todos los métodos, salvo con el método con kernel polinómico de tercer grado, que lo hace cuando $C=15$. Estas tasas iniciales son del 99.9% para el método con kernel gaussiano, le sigue el lineal con una tasa de 99.7%, después el KDCV con $n=3$ con un éxito del 99.6% y por último el KDCV con $n=2$ con una tasa del 99.4%.

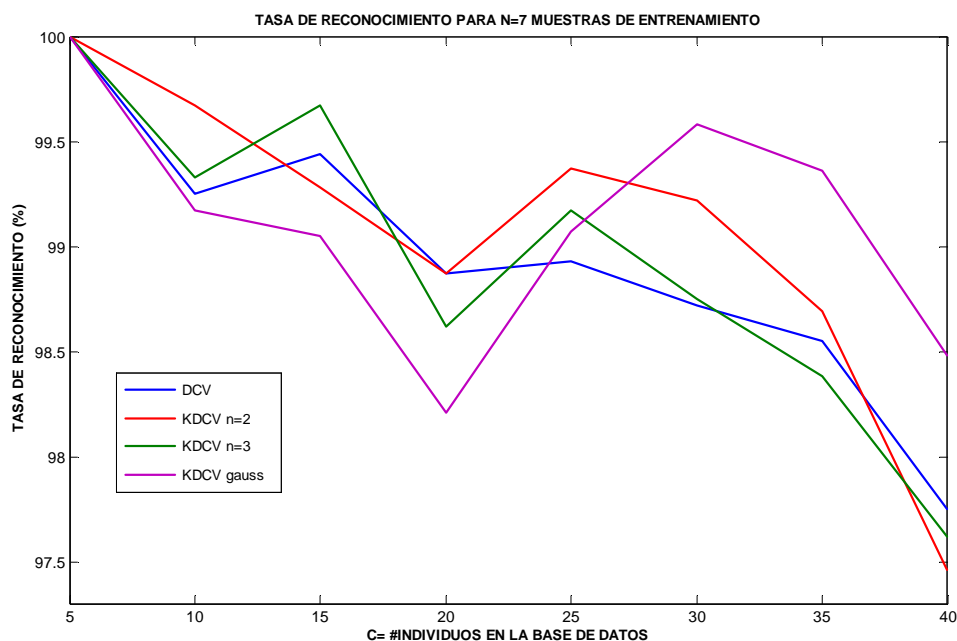
Cuando tenemos un mayor número de sujetos, 40, el método KDCV con kernel gaussiano es el que se sigue manteniendo superior con una tasa del 96.4%. Pero en este caso, el método lineal ha sufrido una mayor reducción del rendimiento que el KDCV con $n=2$ y es éste el que ocupa el segundo lugar con una tasa de reconocimiento del 96.31%. A continuación se sitúan los métodos lineal y con kernel polinómico de tercer grado, que obtienen tasas de 95.81% y 95.56% respectivamente.

Cabe destacar el hecho de que se produzca un máximo local destacado cuando trabajamos en un conjunto de 15 sujetos. Todos los métodos, con este número de individuos, elevan su tasa aunque solamente supera la tasa inicial que se produce con 5 clases el método KDCV con $n=3$. Este método obtiene en este caso una tasa de 99.9%.

Concluyendo, desde que se evalúan los métodos en una base de datos con 5 personas hasta que el tamaño de la misma es de 40 individuos, los diferentes algoritmos han decrementado sus tasas en 4.04% el KDCV con $n=3$; 3.89% el lineal; 3.5% el de kernel gaussiano y 3.09% el de kernel polinómico de segundo grado. Es por tanto el KDCV con $n=3$ el que produce la menor tasa cuando evaluamos con el mayor valor de C y el que, en todo el intervalo de C , más disminuye su rendimiento.

GRAFICA 4.5

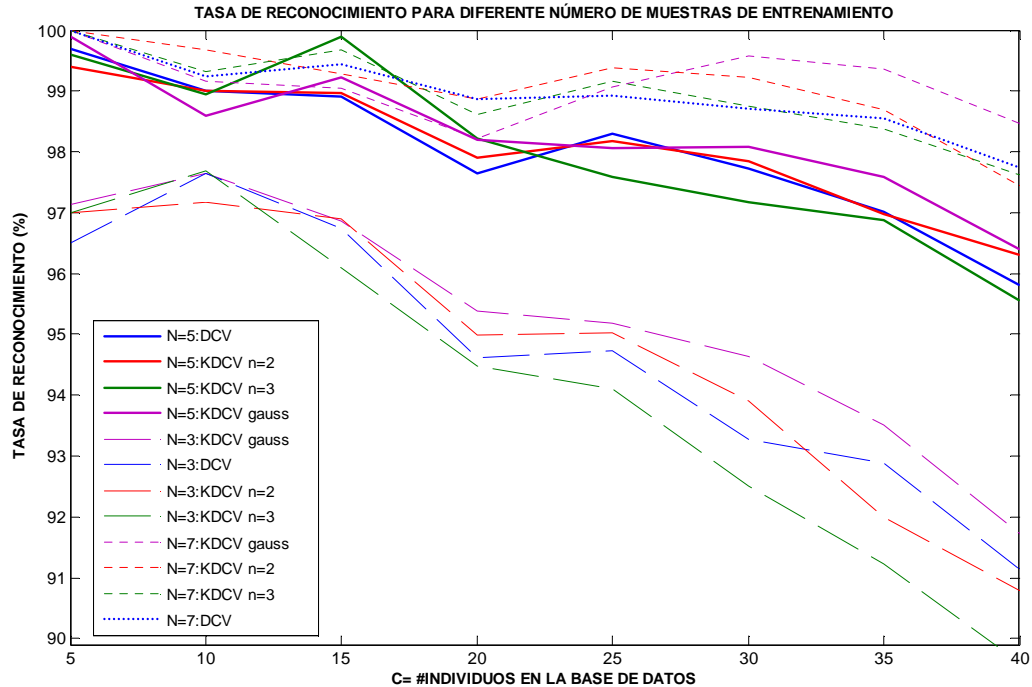
TASA DE RECONOCIMIENTO PAR LA BASE DE DATOS ORL CON N=7 EN FUNCIÓN DEL NÚMERO DE INDIVIDUOS



Para el caso en el que entrenamos con 7 imágenes por individuo, las tasas también decrecen cuando aumenta el número de sujetos pero de manera menos homogénea. Es decir, las tasas comienzan para $C=5$ con un 100% de reconocimiento con todos los métodos y a continuación van decreciendo para remontar en $C=10$ y producir un máximo local en $C=15$. Ocurre otro mínimo local el $C=20$ y otro máximo local en $C=25$ para los métodos de kernel polinómico y en $C=30$ para el de kernel gaussiano. A partir de este momento el método KDCV con kernel gaussiano, como viene siendo habitual, ofrece las mejores tasas llegando a casi un 98.5% para $C=40$. Después se encuentra el método KDCV con $n=2$, aunque para $C=40$ es el que ofrece una tasa inferior. DCV y KDCV con $n=3$ se mantienen similares y algo inferiores a los otros dos métodos, terminando, para el mayor número de individuos, con tasas de 96.75% y 97.62% respectivamente.

En todo el intervalo desde que evaluamos con 5 individuos hasta que realizamos el experimento con 40, los métodos con kernels polinómicos son los que más reducen su rendimiento, un 2.54% el KDCV con $n=2$, y un 2.38% el KDCV con $n=3$. A continuación se sitúa el lineal, que lo hace en un 2.25% y por último, el que menos baja su tasa de rendimiento es el de kernel gaussiano, en un 1.52%.

No obstante, las tasas que obtenemos son muy buenas y su degradación con el tamaño de la base de datos es muy ligera, obteniendo tasas en todo momento de más de 97.4%.

GRAFICA 4.6**TASA DE RECONOCIMIENTO PAR LA BASE DE DATOS ORL EN FUNCIÓN DEL NÚMERO DE INDIVIDUOS, CON DIFERENTE NÚMERO DE MUESTRAS DE ENTRENAMIENTO**

Cuando entrenamos los métodos con un número pequeño de muestras, 3 por sujeto, podemos ver que conseguimos tasas menores que en el resto de los casos. Parece lógico esperar este resultado, ya que no habrá suficiente espacio para obtener los vectores de proyección óptimos al construir nuestra máquina con una cantidad de información menor. Ocurre que el espacio nulo de la matriz de dispersión dentro de la clase, S_w , cuando entrenamos con 3 muestras es más pequeño ya que M en este caso sería 15, 30, 45,..., 120 en función del número de individuos con el que componamos la base de datos.

Como podemos observar, obtenemos mejores tasas en todos los casos cuando el número de sujetos es menor, hecho que ya habíamos comentado con anterioridad a la vista de los resultados para cada N_{train} . Para $N_{\text{train}} = 3$ obtenemos las mayores tasas para todos los métodos cuando $C=10$, mientras que para $N_{\text{train}} = 5$ y 7 esto ocurre cuando

$C=5$; llegando incluso a una tasa del 100% para todos los métodos cuando entrenamos 7 muestras.

En general, las tasas que obtenemos son bastante buenas, aunque se degradan algo cuando aumentamos la base de datos, las tasas que obtenemos con $C=40$ son de entre 95.5% y 98.5% cuando entrenamos con $N_{\text{train}}=5$ y $N_{\text{train}}=7$ y algo menores, entorno a 91%, cuando $N_{\text{train}}=3$. Observamos que cuando entrenamos con 5 y 7 muestras, las tasas que obtenemos son más próximas que en el caso de $N_{\text{train}}=3$, caso en el que vemos que el reconocimiento empeora bastante con respecto a resultados anteriores. Se mantiene un comportamiento similar entre el entrenamiento con 5 imágenes y con 7, sobre todo, cuando tenemos una base de datos pequeña, es decir, $C < 20$. A partir de este momento los resultados empiezan a diferir más, siendo mejores las tasas de reconocimiento cuando $N_{\text{train}}=7$.

El método KDCV con kernel gaussiano produce los mejores resultados. Cuando $N=7$, se comporta mejor que el resto de los métodos para todos los posibles números de individuos en la base de datos. Le sigue el método KDCV con kernel polinómico de grado 2, aunque le supera el método DCV cuando $C \geq 35$. Finalmente, el método KDCV con kernel polinómico de grado 3 es el que ofrece tasas más bajas para $C \geq 10$.

4.4. Robustez de los métodos DCV y KDCV frente a cambios en la iluminación

En este caso, se elige la base de datos Yale [58] ya que nos encontramos imágenes en su mayoría con luz centrada a excepción de dos de ellas, que son con iluminación derecha e izquierda respectivamente. La base de datos dispone de 11 imágenes por individuo y un total de 15 individuos. El tamaño de las imágenes es de $320 * 243$ píxeles, de manera que no nos es necesario reducir la dimensión porque no nos limita su tratamiento en *matlab*. Además, no realizamos ningún tipo de preprocesado a las imágenes, ni la extracción de color ya que las muestras están en escala de gris.

Hemos evaluado la tasa de reconocimiento en esta base previamente con el método *leave one out*. Dado que este procedimiento fue utilizado para test y que hubo solamente dos muestras con estas condiciones de iluminación por clase, estas imágenes de iluminación inusual se comportan como datos anómalos (es decir, estas imágenes distan mucho de las muestras utilizadas para el entrenamiento). Serían de esperar mejores resultados si hubiera más de una muestra con estas condiciones de iluminación.

Las direcciones de proyección minimizan el total de la dispersión dentro de las clases utilizadas. Una parte importante del total de la dispersión dentro de la clase se produjo por los casos de iluminación no ambiente en todas las clases. Esta variación debida a las condiciones lumínicas parece producir desviaciones similares de todas las clases respecto de la clase media. Por lo tanto, es razonable pensar que la proyección resultante reduce la variación debida a la iluminación en todas las clases, incluso en las clases en las que tales variaciones no aparecen en el conjunto entrenamiento.

Para probar la robustez de estos métodos respecto a esta variación, tomamos 7 imágenes aleatorias de cada individuo para entrenamiento, entre las que no se encuentran las imágenes con iluminación lateral derecha e izquierda. Con las cuatro

restantes hacemos la evaluación, es decir, con las dos imágenes de iluminación lateral hacemos un test y con las otras dos restantes, que no están sujetas a iluminación lateral, hacemos otro test. Repetimos este procedimiento 10 veces y se hallaron dos tasas de reconocimiento (una para el caso de iluminación lateral y otra para el caso de iluminación natural) que fueron calculadas como la media de las tasas de reconocimiento de cada ejecución para cada caso. Una de las tasas se corresponde a la situación en la que no tenemos cambio en la iluminación mientras que la otra se corresponde con la tasa de reconocimiento cuando introducimos cambios severos en la iluminación. Los resultados hallados se pueden ver en la Tabla 4.7.

TABLA 4.7

TASA DE RECONOCIMIENTO CON CAMBIO DE ILUMINACIÓN PARA LA BASE DE DATOS YALE

Condiciones de iluminación	Tasas de Reconocimiento y Desviaciones Típicas							
	DCV		Kernel DCV n=2		Kernel DCV n=3		Kernel Gaussiano	
	tasa %	σ	tasa %	σ	tasa %	σ	tasa %	σ
SIN CAMBIO DE LUZ	96.33	5.31	96.33	5.31	95.67	6.67	96.33	5.54
CON CAMBIO DE LUZ	41.00	8.61	33.33	7.37	25.67	3.16	40.33	5.54

Como podemos observar, se produce una degradación muy pronunciada en las tasas de reconocimiento cuando evaluamos los métodos con muestras que contienen una iluminación izquierda y derecha, diferente a la usada para entrenamiento. Para todos los métodos este descenso del rendimiento es muy notable. En todos los casos, cuando hacemos el test con imágenes de los individuos con condiciones de iluminación centrada o ambiente, igual que la usada en el entrenamiento, conseguimos tasas de reconocimiento muy altas, todas ellas similares de entorno a un 96%. Sin embargo, cuando aplicamos una variación en este factor objeto de estudio, todas estas tasas descienden abruptamente, aunque unas más que otras. En el caso del método DCV, la tasa disminuye más de la mitad, como en el caso del KDCV con kernel gaussiano,

siendo el valor de ambas de entorno al 40%. En los otros dos casos, este descenso de la tasa de reconocimiento es mucho mayor, siendo en el caso del método KDCV con kernel polinómico de $n=3$ de 33.3% y en el caso del KDCV con kernel polinómico de $n=2$ apenas 25%.

Podemos afirmar entonces que estos métodos son, extremadamente sensibles a los cambios de iluminación, comportándose el DCV lineal y el KDCV gaussiano algo más robustos, y siendo el KDCV con kernel polinómico de $n=3$ el menos robusto frente a esta variación.

4.5. Robustez de los métodos DCV y KDCV frente a oclusiones parciales

Para evaluar el impacto de las oclusiones en el reconocimiento facial, realizamos un experimento con la base de datos de Aleix Martínez y Robert Benavente (AR) [59]. Esta base de datos contiene más de 4000 imágenes a color; 126 individuos y 26 imágenes por individuo. Estas imágenes contienen oclusiones. Las siete primeras fueron tomadas 2 semanas antes que las siete últimas y en ambas sesiones se tomaron imágenes con oclusiones parciales como bufandas o pañuelos y gafas de sol, que se combinan con iluminación lateral. En total tenemos 14 imágenes sin oclusiones y 12 con oclusiones parciales.

En nuestra tarea de analizar la robustez de los métodos DCV y KDCV frente a estas oclusiones parciales, escogemos 50 individuos de la base de datos. Como el tamaño de las muestras, $576 * 768$ píxeles, es elevado, cuando aplicamos los métodos con *matlab* nos encontramos con serios problemas de memoria y necesitamos reducir el tamaño para poder sacar los vectores comunes. Hacemos un diezmado de factor 2 en todas las imágenes de la base de datos y finalmente su tamaño es de $288 * 384$ píxeles. A continuación convertimos las imágenes a escala de gris. Realizamos el entrenamiento con 7 imágenes escogidas aleatoriamente del conjunto de imágenes que no contienen oclusiones y posteriormente hacemos dos evaluaciones, una de ellas con las 7 imágenes restantes del conjunto de fotos sin oclusiones, que producirá la tasa de reconocimiento sin oclusiones; y la otra con 7 imágenes escogidas al azar del conjunto de imágenes que contienen oclusiones. Este procedimiento lo repetimos hasta 10 veces y las tasas finales de reconocimiento las calculamos como el promedio de las halladas en cada ejecución.

Los resultados obtenidos se muestran en la Tabla 4.8

TABLA 4.8**TASA DE RECONOCIMIENTO CON Y SIN OCLUSIONES PARA LA BASE DE DATOS AR**

Presencia de Oclusiones	Tasas de Reconocimiento (%) y Desviaciones Típicas							
	DCV		Kernel DCV n=2		Kernel DCV n=3		Kernel Gaussiano	
	tasa %	σ	tasa %	σ	tasa %	σ	tasa %	σ
SIN OCLUSIONES	98.03	2.21	99.14	2.98	99.14	2.67	98.23	2.43
CON OCLUSIONES	53.28	10.49	60.57	3.06	61.41	3.16	55.51	9.60

Como podemos comprobar, cuando realizamos el reconocimiento sin oclusiones, las tasas que obtenemos son realmente altas. Conseguimos las mayores tasas para los métodos KDCV con kernels polinómicos (99.14 %). La menor tasa la proporciona el método lineal, que produce un rendimiento del 98%. Sin embargo, cuando realizamos el test con imágenes que contienen oclusiones, las tasas de acierto disminuyen considerablemente. En el caso del método KDCV con kernel polinómico de grado 2, la tasa cae a un 60 % aproximadamente, al igual que ocurre con el KDCV de $n=3$. También decrecen las tasas de reconocimiento de los métodos DCV y KDCV con kernel gaussiano a un 53.3 % y un 55.5 % respectivamente. El método más sensible si cabe es el método lineal, que ve disminuida su tasa en casi un 45 %. El resto de métodos lo hace propiamente, entorno a un 40 %. No obstante se mantienen algo más fiables los métodos KDCV con kernel polinómico.

Podemos afirmar a la vista de los resultados, que existe una debilidad en esta técnica de reconocimiento facial basada en vectores comunes: el alto grado de sensibilidad frente a oclusiones parciales. Como los métodos hacen un análisis basado en la información que proporciona cada píxel, cuándo a cierta cantidad de píxeles se les priva de información, es lógico que estos métodos ofrezcan peores resultados.

4.6. Robustez de los métodos DCV y KDCV frente al ruido

Otra de las dificultades con las que nos podemos encontrar a la hora de tratar imágenes es el ruido, que puede producirse en la captura de las imágenes, la transmisión o el procesamiento de las mismas. Este factor puede hacer que sea más difícil extraer la información deseada, es decir, la identidad de la persona. Precisamente, vamos a evaluar este impacto en los métodos de reconocimiento facial basados en vectores comunes.

El ruido es la información no deseada que contamina la imagen.

$$g(x,y) = f(x,y) + r(x,y)$$

donde $g(x,y)$ es la imagen resultante, $f(x,y)$ la original y $r(x,y)$ el ruido. El origen puede estar tanto en el proceso de adquisición de la imagen (errores en los sensores), como en la transmisión (debido a interferencias en el canal de transmisión). Existen modelos de ruido, según las funciones de densidad de probabilidad que sigan sus intensidades $r(x,y)$:

- Gaussiano \rightarrow o normal, modela el ruido producido por los circuitos electrónicos o ruido de los sensores por falta de iluminación y/o altas temperaturas. La distribución del ruido se asemeja a una distribución gaussiana de una determinada media y varianza
- Impulsivo (Sal y Pimienta) \rightarrow o aleatorio, se produce normalmente en la cuantificación que se realiza en el proceso de digitalización. El número de píxeles afectados por el ruido y su intensidad son las variables aleatorias independientes entre sí.
- Uniforme \rightarrow Toma valores en un determinado intervalo de forma equiprobable. Se da en un menor número de situaciones reales; por lo que no será considerado en nuestro estudio.

- Etc.

Desde el inicio de los tiempos, el hombre se ha topado con la necesidad de comunicarse de manera correcta, pero se ha visto afectado por factores externos a él. En el caso específico de las imágenes, tema a tratar en este proyecto, los factores que afectan a las imágenes en el momento de obtenerlas son el movimiento del objeto o cuerpo a ser tomado, o que se encuentre fuera de foco; y en el momento de la transmisión de esta imagen puede ser afectada por la turbulencia atmosférica y el ruido Gaussiano.

Se eligió el Ruido Gaussiano, ya que es el ruido más común en transmisiones de radio y satelitales. Además se seleccionó el Ruido Sal y Pimienta, ya que se aprecia una imagen difícil de restaurar con salpicaduras y variaciones de tonos drásticos.

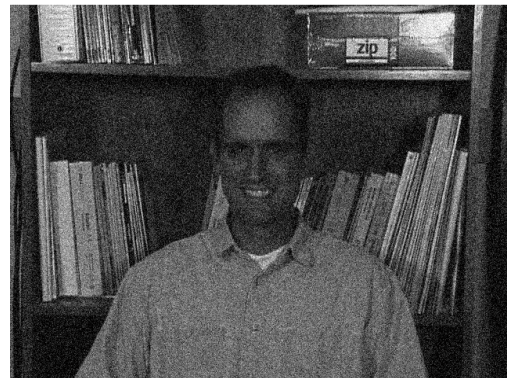
Para analizar la robustez de los métodos DCV y KDCV frente al ruido, vamos a evaluarlos en la base de datos Georgia Tech [61]. Usamos esta base de datos de caras transformándolas previamente a escala de gris. Como las imágenes tienen un tamaño de $640 * 480$ píxeles, nos vemos obligados a reducirlas a la mitad por necesidades del programa utilizado para implementar los métodos, *matlab*. Esta base de datos contiene 750 imágenes, con un total de 50 individuos. No contiene oclusiones ni cambios en la iluminación, de manera que trataremos de estudiar la influencia del ruido sin que ninguno de estos factores afecte al reconocimiento.

4.6.1. Robustez de los métodos DCV y KDCV frente al Ruido Gaussiano

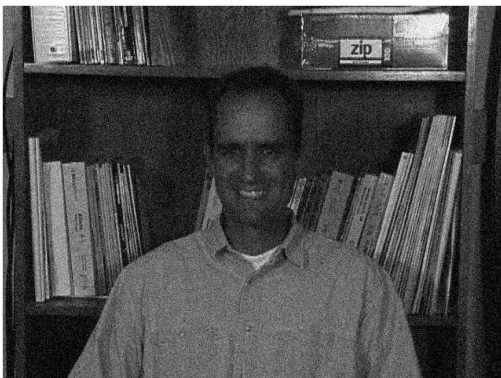
Escogemos aleatoriamente, de las 15 muestras por individuo que hay, 8 imágenes para entrenar y las 7 restantes para test. Realizamos el entrenamiento con las imágenes sin más procesado que el citado anteriormente, que consiste en disminuir su tamaño (es de $320 * 240$ píxeles) y pasarlas a escala de gris. A continuación realizamos diversos test añadiendo ruido Gaussiano. Lo hacemos primero sin añadir ruido y después con diferentes varianzas de ruido: $v=0.000$, $v=0.005$, $v=0.01$, $v=0.015$, $v=0.02$, $v=0.025$, $v=0.05$, $v=0.075$, $v=0.1$, $v=0.15$, $v=0.2$, $v=0.4$, $v=0.5$, $v=0.6$ y $v=0.7$ $v=1$, $v=3$, $v=5$ y $v=40$. Este proceso lo realizamos 10 veces, eligiendo siempre las muestras de entrenamiento para cada ejecución aleatoriamente. La tasa de reconocimiento se calcula como la media de las tasas obtenidas en cada realización. A continuación, en la Figura 4.1 podemos observar una muestra de la base de datos GT transformada a escala de gris y con el ruido Gaussiano agregado para la realización de nuestro análisis.



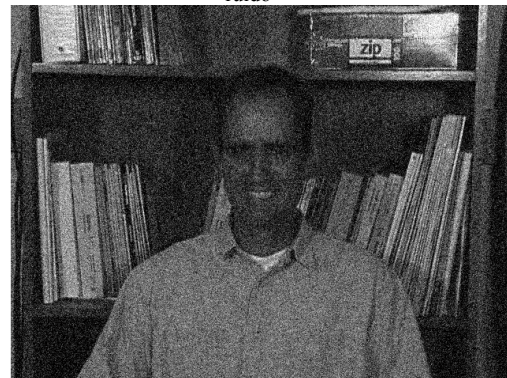
Sin ruido



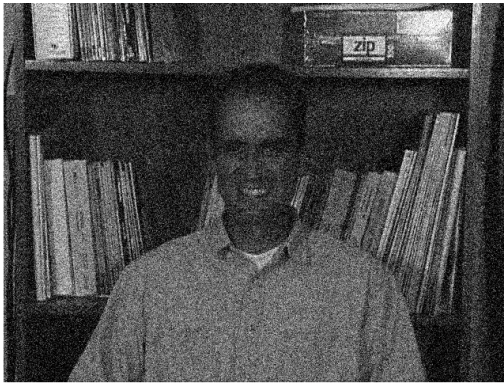
$\text{var}_{\text{ruido}}=0.010$



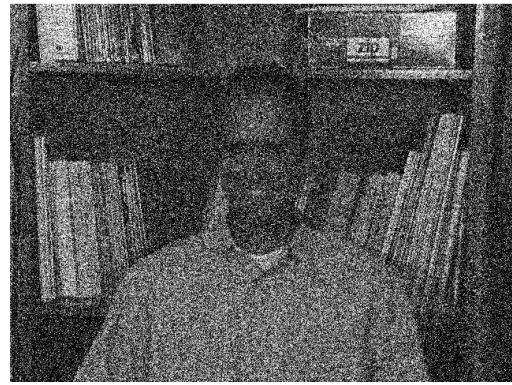
$\text{var}_{\text{ruido}}=0.005$



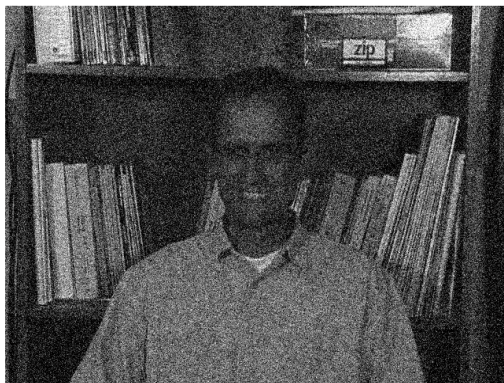
$\text{var}_{\text{ruido}}=0.015$



$\text{var}_{\text{ruido}}=0.020$



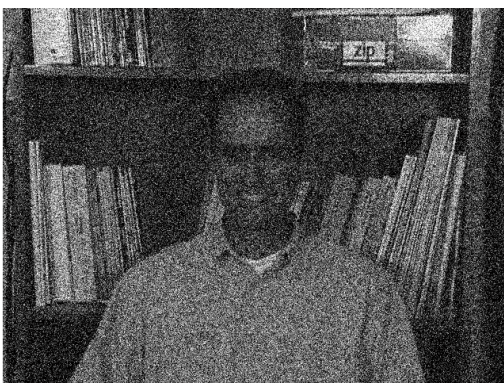
$\text{var}_{\text{ruido}}=0.1$



$\text{var}_{\text{ruido}}=0.025$



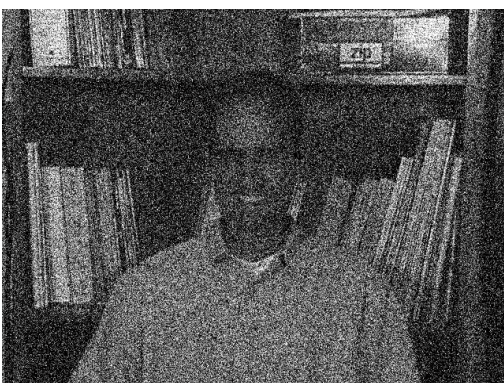
$\text{var}_{\text{ruido}}=0.15$



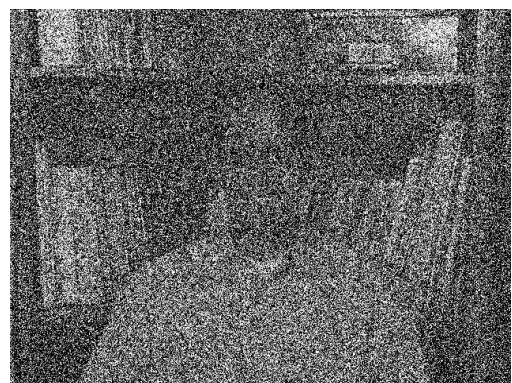
$\text{var}_{\text{ruido}}=0.05$



$\text{var}_{\text{ruido}}=0.2$



$\text{var}_{\text{ruido}}=0.075$



$\text{var}_{\text{ruido}}=0.4$

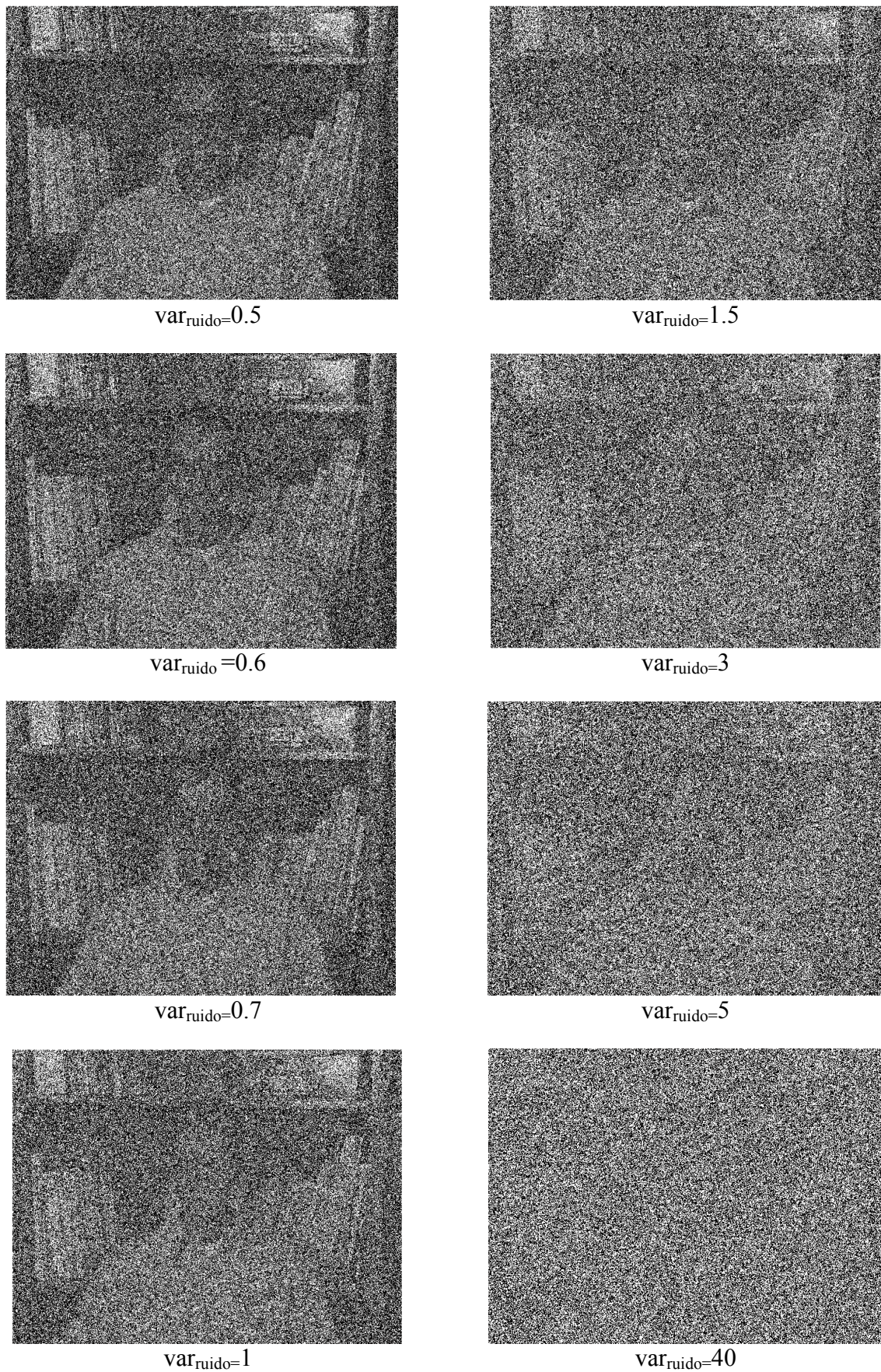


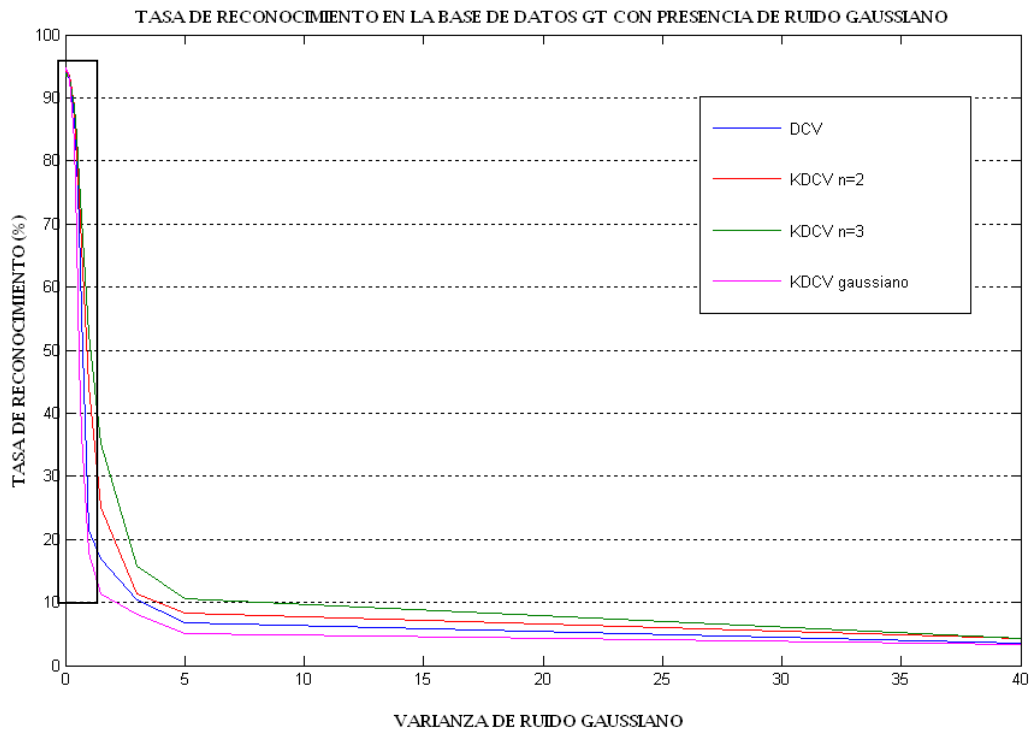
Figura 4.1: Un ejemplo de imagen de la base de datos GT al que le sumamos ruido Gaussiano de diferente varianza

Los resultados obtenidos se reflejan en la Tabla 4.9 y su correspondiente representación en las Gráfica 4.7, que se ha mostrado con más detalle en las Gráficas 4.8 y 4.9.

TABLA 4.9

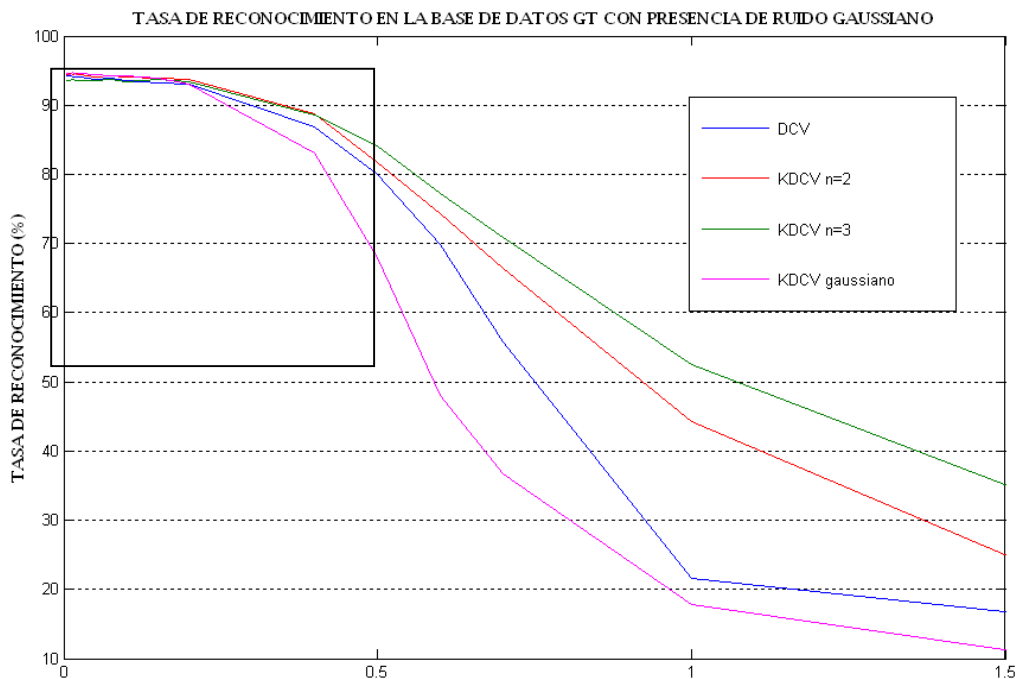
TASA DE RECONOCIMIENTO EN LA BASE DE DATOS GT CON PRESENCIA DE RUIDO GAUSSIANO

Ruido Gaussiano	Tasas de Reconocimiento (%) y Desviaciones Típicas							
	DCV		Kernel DCV n=2		Kernel DCV n=3		Kernel Gaussiano	
	Tasa%	σ	tasa%	σ	tasa%	σ	tasa%	σ
v=0.000	94.23	1.68	94.49	1.79	93.77	1.49	94.57	1.92
v=0.005	94.26	1.54	94.43	1.87	93.60	1.64	94.66	1.86
v=0.010	94.20	1.67	94.57	1.77	93.60	1.59	94.37	1.95
v=0.015	94.14	1.77	94.40	2.04	93.71	1.50	94.69	1.86
v=0.020	94.03	1.82	94.57	1.71	93.60	1.53	94.49	1.83
v=0.025	94.14	1.73	94.34	1.96	93.60	1.53	94.54	1.92
v=0.050	93.80	2.07	94.14	1.92	93.49	1.27	94.46	1.99
v=0.075	93.83	2.33	94.26	2.14	93.71	1.19	94.29	2.08
v=0.100	93.57	2.51	94.09	2.11	93.43	1.38	94.31	2.20
v=0.150	93.26	2.67	93.89	2.12	93.71	1.05	93.89	2.29
v=0.200	93.09	2.73	93.69	2.50	93.31	1.08	93.09	3.29
v=0.400	86.86	7.63	88.69	5.63	88.63	2.72	83.14	9.95
v=0.500	80.00	11.24	81.71	9.23	84.11	4.27	67.94	11.92
v=0.600	69.80	13.83	74.26	12.26	77.20	9.86	47.97	10.60
v=0.700	55.77	12.31	66.49	12.61	70.94	10.28	36.77	8.44
v=1.000	21.56	3.68	44.34	10.12	52.57	10.13	17.77	3.56
v=1.500	16.83	3.02	25.05	4.51	35.20	8.67	11.37	2.17
v=3.000	10.36	2.31	11.26	2.16	15.71	2.39	8.03	1.62
v=5.000	6.83	1.22	8.31	2.16	10.54	2.11	5.11	2.29
v=40.00	3.51	1.03	4.34	1.48	4.34	1.82	3.31	0.82

GRÁFICA 4.7**TASA DE RECONOCIMIENTO EN LA BASE DE DATOS GT CON PRESENCIA DE RUIDO GAUSSIANO**

Como es lógico esperar, las tasas de reconocimiento disminuyen cuando el ruido gaussiano que añadimos a las imágenes aumenta. Cuando no añadimos ruido a las muestras, las tasas de reconocimiento son bastante buenas, siendo similares todas ellas en torno a un 94%.

A la vista de la Gráfica 4.7, podemos apreciar que la mayor variación del porcentaje de aciertos que producen todos los métodos, se da entre una varianza de ruido nula y una varianza de ruido de 5. A partir de entonces, las tasas se mantienen muy bajas pero algo más constantes, situándose en todo momento inferiores al 10%. Entre una varianza de 5 y 40, los métodos tan sólo disminuyen sus tasas entre un 2% y un 6% aproximadamente, quedando para este último valor de ruido, un éxito de clasificación en el mejor caso, del 4.34% (KDCV, $n=3$), y 3.31% en el peor (KDCV con kernel gaussiano). Para una varianza de 5, podemos ver en las imágenes mostradas con ruido, que la imagen es indistinguible para el ojo humano. Sin embargo, los métodos nos proporcionan tasas mayores que la que nos proporcionaría el azar, 0.2.

GRÁFICA 4.8**DETALLE DE LA TASA DE RECONOCIMIENTO EN LA BASE DE DATOS GT CON PRESENCIA DE RUIDO GAUSSIANO**

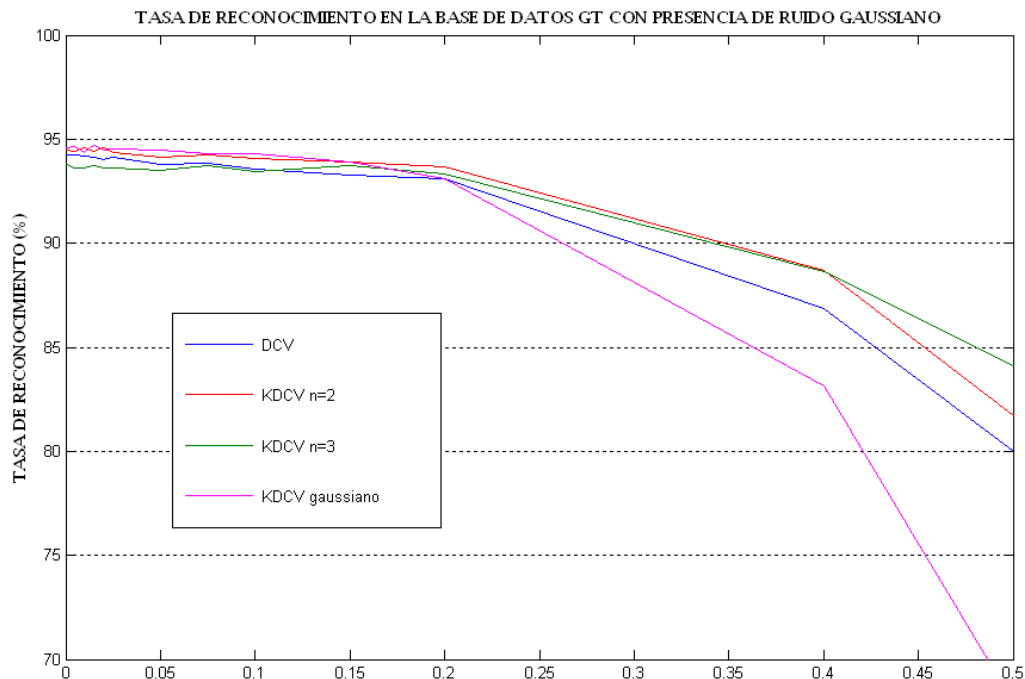
Si observamos con más detalle las tasas obtenidas con una varianza de ruido Gaussiano de hasta 1.5, apreciamos una mayor diferencia de la capacidad de generalización de los métodos. Nos centramos en el tramo comprendido entre una varianza de 0.4 y 1.5. Desde que comienza este tramo, el método que mejor clasifica datos desconocidos es el método KDCV con kernel polinómico de tercer grado. Le sigue el método KDCV con $n=2$, después el lineal y por último el de kernel gaussiano. Para una varianza de 0.4, vemos en las imágenes mostradas como ejemplo que la presencia de ruido es importante y la imagen está muy distorsionada. En este caso, el éxito de los métodos KDCV con kernel polinómico es similar, del 88.6%. Reconocemos que esta tasa es muy elevada dada la dificultad que nos supone visualmente reconocer a la persona de la imagen. Estos métodos descienden de manera muy parecida, haciéndolo algo más el de grado 2. Éste pasa a obtener una tasa del 25% (empeora un 66% aproximadamente) cuando a varianza es 1.5, mientras que el KDCV obtiene un 10% más, decrementando su tasa un 56%.

Con menores tasas que los KDCV con kernel polinómico y con un descenso más abrupto de las mismas, se sitúan los métodos DCV y KDCV con kernel gaussiano.

Estos métodos comienzan el intervalo con un porcentaje de éxito del 88.9% y 83.1% respectivamente y lo terminan con tasas del 16.8% y 11.3%. Es decir, empeoran considerablemente, lo hacen ambos en un 72%.

GRÁFICA 4.9

DETALLE DE LA TASA DE RECONOCIMIENTO EN LA BASE DE DATOS GT CON PRESENCIA DE RUIDO GAUSSIANO



Por último, vemos con algo más de detalle el comportamiento que se sigue cuando tenemos varianzas de ruido gaussiano más pequeñas. Dividimos el análisis en dos subintervalos, el que comprende hasta una varianza de 0.2 y el que abarca desde este punto hasta una varianza de 0.4.

En el segundo subintervalo apreciamos un dominio del método KDCV con $n=2$; situándose con una tasa superior a las del resto de métodos. En este tramo se produce una alteración con respecto al comentado anteriormente, y es que el método KDCV con $n=3$ pierde la condición de ser el más exitoso que mantenía cuando estábamos en presencia de ruido mayor. Aunque KDCV con $n=2$ y $n=3$ difieren muy poco. Para $v_{\text{ruido}}=0.2$, las tasas de todos los métodos son muy parecidas, del 93% para los métodos

lineal y de kernel gaussiano y del 93.7% y 93.3% para los polinómicos de $n=2$ y $n=3$ respectivamente. Por tanto, desde que tenemos una varianza de 0.2 hasta que la tenemos de 0.4, el método que más ha visto reducida su capacidad de generalización es el de kernel gaussiano, que lo hace en un 6.6%. Le sigue el lineal con un 6.2% y algo menos los polinómicos con un 5% ($n=2$) y 4.7% ($n=3$).

En el primer tramo, hasta que tenemos una varianza de 0.2, los métodos siguen una tendencia de ligero descenso. Todos ellos ven decrementada su tasa de manera muy sutil, apenas un 1.2% en el peor de los casos. Para las tasas más pequeñas, se mantiene superiores los métodos de kernel gaussiano y de kernel polinómico de grado 2. El que obtiene una tasa inferior al principio es el de $k=3$, que proporciona un 93.6% cuando aplicamos $\text{var}_{\text{ruido}}=0.2$. Después pasa a ser inferior el lineal y a partir de esta varianza de ruido, es el método de kernel gaussiano el que menor probabilidad de éxito produce.

No obstante, con bajo ruido las tasas de reconocimiento son muy buenas. Sobre todo si comparamos estos resultados con nuestra apreciación del ruido mostrado en las imágenes.

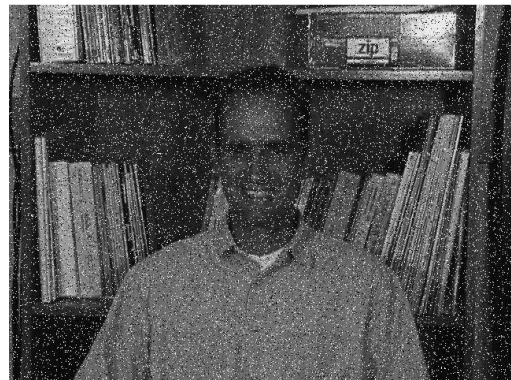
4.6.2. Robustez de los métodos DCV y KDCV frente al Ruido Sal y Pimienta

Realizamos el mismo experimento que el descrito anteriormente con el ruido Gaussiano, Es decir, elegimos aleatoriamente 8 imágenes previamente reescaladas y transformadas a escala de gris para realizar el entrenamiento. A continuación, con las 7 restantes realizamos el test. En primer lugar lo hacemos con las imágenes sin ruido y posteriormente con ruido Sal y Pimienta. Realizamos varias ejecuciones con diferentes densidades de este ruido impulsivo. Éstas tomarán unos valores de 5%, 10%, 15%, 20%, 40%, 50%, 60%, 80% y 100%.

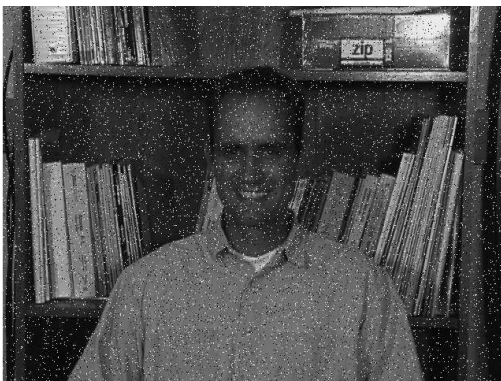
A continuación, en la Figura 4.2 podemos observar una muestra de la base de datos transformada a escala de gris y a la que hemos ido sumando ruido Sal y Pimienta con diferentes densidades.



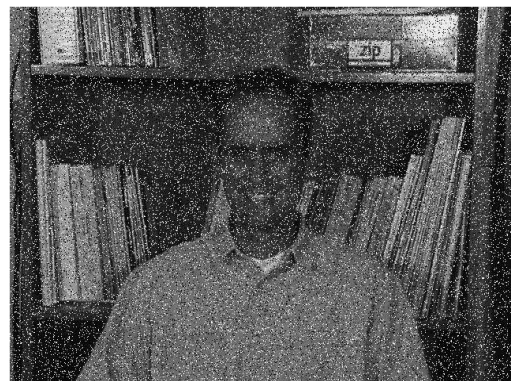
Sin ruido



$d_{\text{ruido}}=10 \%$



$d_{\text{ruido}}=5 \%$



$d_{\text{ruido}}=15 \%$

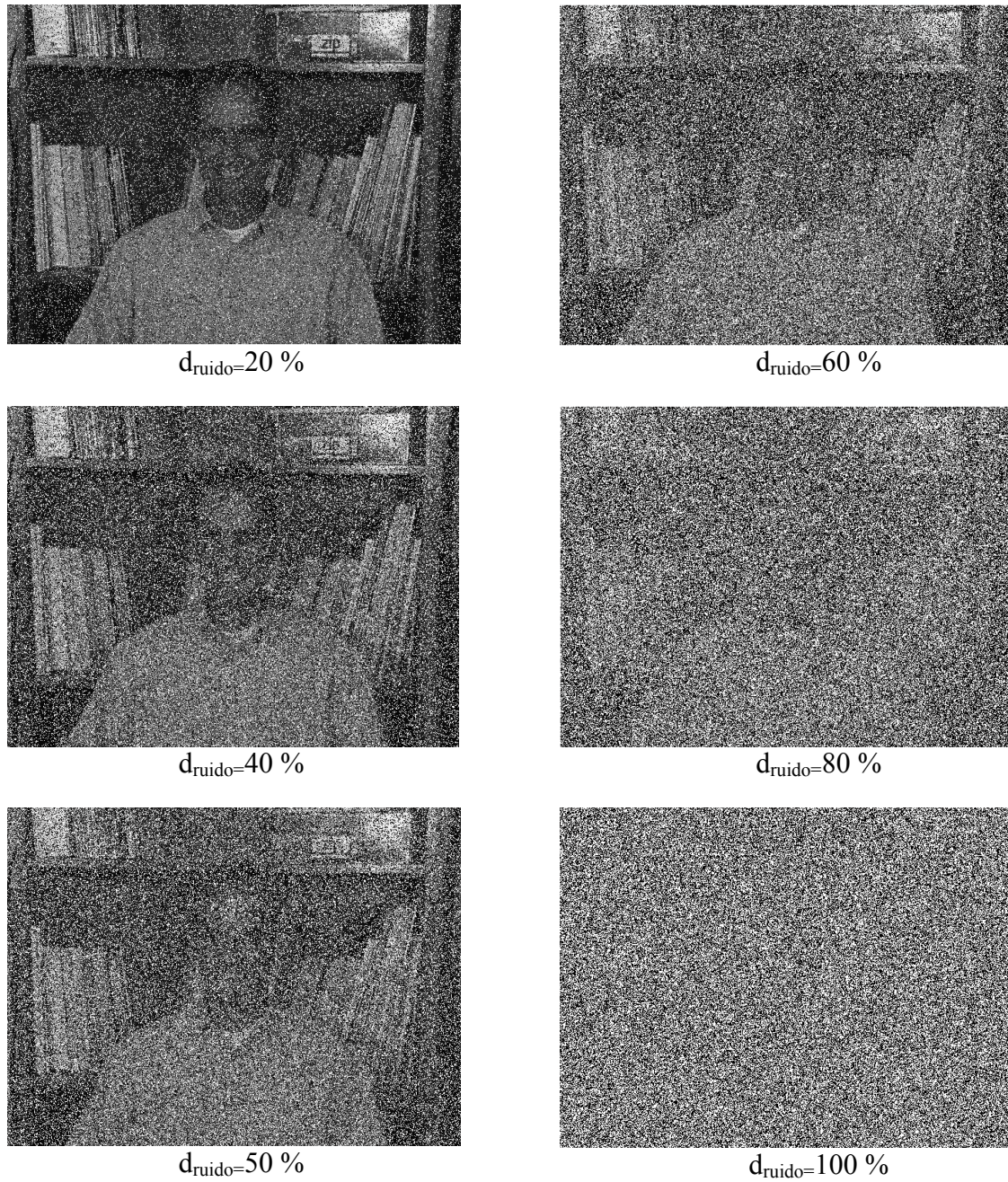


Figura 4.2: Un ejemplo de imagen de la base de datos GT al que le sumamos ruido Sal y Pimienta de diferente densidad

En la Tabla 4.10 se reflejan los resultados obtenidos en nuestro experimento. Además, podemos ver una representación de estas tasas halladas en la Gráfica 4.10 y un detalle de ésta en la Gráfica 4.11.

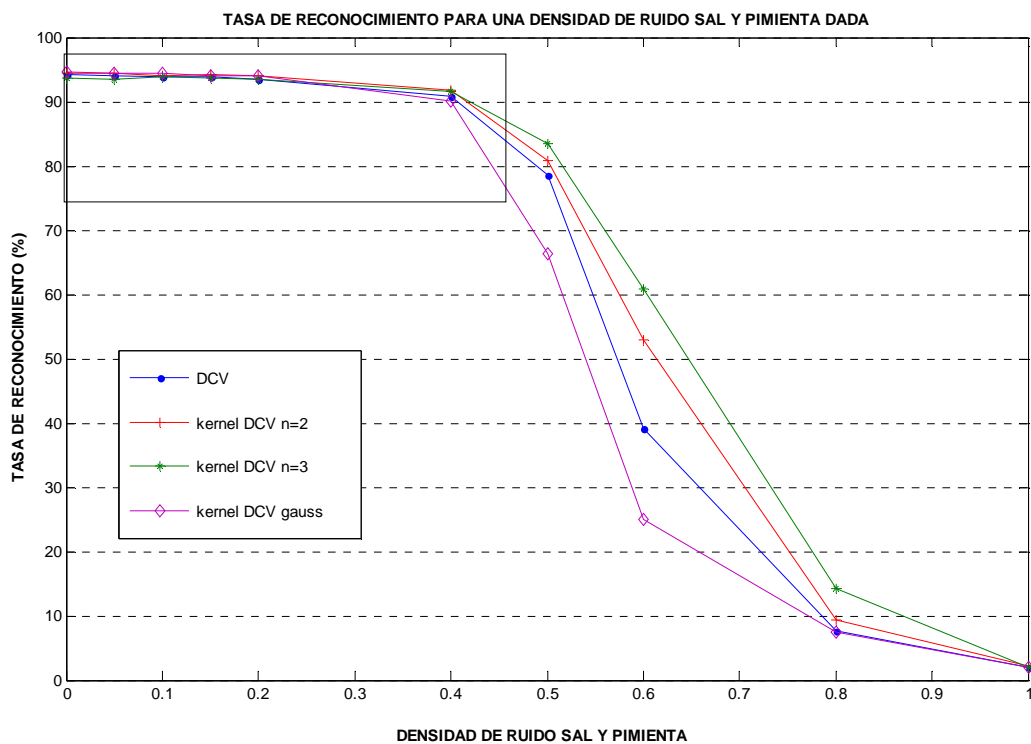
TABLA 4.10

TASA DE RECONOCIMIENTO EN LA BASE DE DATOS GT CON PRESENCIA DE RUIDO SAL Y PIMIENTA

Ruido Sal y Pimienta	Tasas de Reconocimiento (%) y Desviaciones Típicas							
	DCV		Kernel DCV n=2		Kernel DCV n=3		Kernel Gaussiano	
	tasa %	σ	Tasa %	σ	tasa %	σ	tasa %	σ
D=0%	94.23	1.66	94.49	1.79	93.77	1.49	94.57	1.92
d = 5%	94.14	1.80	94.46	1.86	93.54	1.41	94.46	1.96
d = 10%	93.89	2.03	94.14	2.14	93.89	1.30	94.46	1.92
d = 15%	93.86	2.34	94.29	2.10	93.66	1.42	94.06	2.22
d = 20%	93.54	2.51	94.03	2.40	93.43	1.28	94.06	2.22
d = 40%	90.89	4.39	91.77	3.42	91.66	2.09	90.09	5.97
d = 50%	78.60	11.65	80.91	9.64	83.40	7.25	66.29	13.37
d = 60%	39.06	7.69	52.97	10.83	60.80	1.98	24.91	5.92
d = 80%	7.63	1.76	9.34	2.02	14.23	2.06	7.43	1.23
d= 100%	1.94	0.22	2.09	0.40	2.06	0.24	2.06	0.32

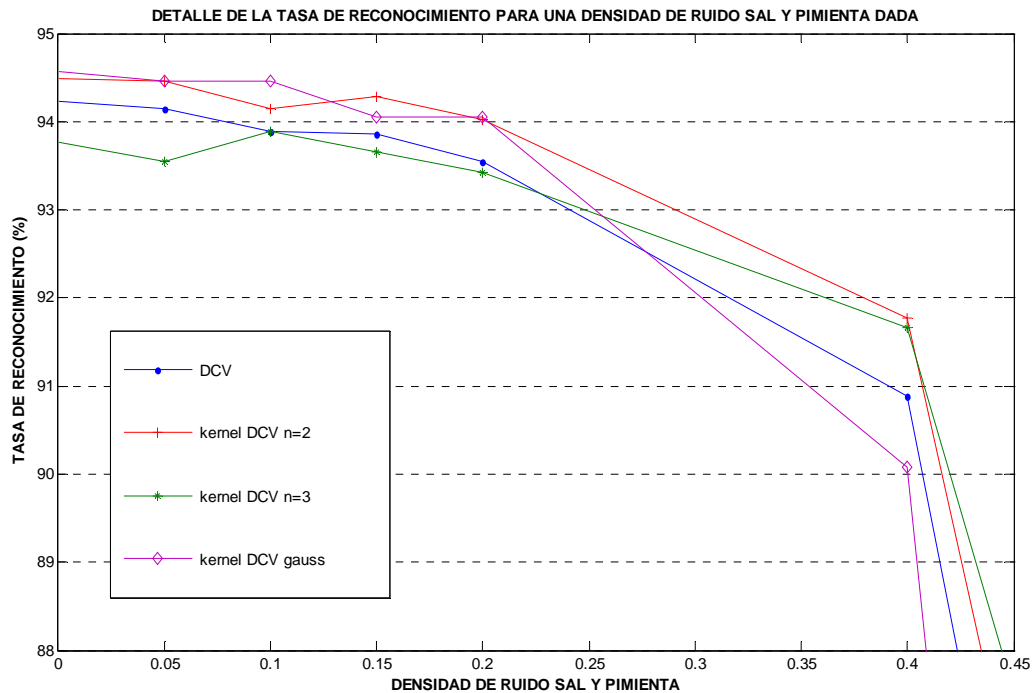
GRÁFICA 4.10

TASA DE RECONOCIMIENTO EN LA BASE DE DATOS GT CON PRESENCIA DE RUIDO SAL Y PIMIENTA



GRÁFICA 4.11

DETALLE DE LA TASA DE RECONOCIMIENTO EN LA BASE DE DATOS GT CON PRESENCIA DE RUIDO SAL Y PIMIENTA



A la vista de los resultados hallados, podemos dividir nuestro análisis en cuatro partes según los tramos que podemos apreciar en las gráficas. Estos tramos vienen delimitados por los valores de densidades de ruido Sal y Pimienta de 20%, 40%, 80% 100%. Es decir, el tramo primero corresponde a las tasas halladas cuando agregamos a las imágenes ruido Sal y Pimienta de densidad de hasta el 20%. El segundo se relaciona con las densidades del 20% al 40%, el tercero del 40% al 80% y finalmente, el cuarto del 80% al 100%. Una densidad del 100% significa que el ruido en la imagen es total, es decir, añadimos ruido Sal y Pimienta a todos los píxeles de la imagen quedando ésta con ruido total y anulando toda componente de la imagen original.

Para estos valores de ruido comprendidos en el primer tramo vemos que las tasas de reconocimiento en todos los casos disminuyen pero muy levemente. Las máximas se obtienen en el caso de los métodos KDCV con kernel gaussiano y con kernel polinómico de grado 2. Ambos producen tasas similares que disminuyen un 0.5 % aproximadamente; es decir, apenas ven reducido su rendimiento. Los otros dos métodos proporcionan peores resultados que los anteriores y bastante similares entre sí, siendo el

DCV ligeramente superior con bajas densidades (hasta 10%) y después el método KDCV con $n=3$. Estos métodos nos dan unas tasas de entorno al 93.5 %, reduciéndose 0.7% y 0.3% aproximadamente respectivamente desde que se evalúan en imágenes sin ruido.

En el segundo tramo, destaca con mejor comportamiento el método KDCV con kernel polinómico de grado 2. El resto mantiene un comportamiento similar, llegando al final del tramo, para el 40% de densidad del ruido Sal y Pimienta, a unas tasas de 91.66% para el método KDCV de $n=3$, 90.89% el lineal y por último, 90.09% el KDCV con kernel gaussiano. En este tramo, el método que más disminuida ve su tasa es el método KDCV gaussiano y lo hace en un 4.5% aproximadamente. El resto de los métodos disminuyen su rendimiento en menor medida, siendo el menos variable el KDCV con $n=3$, cuya tasa disminuye un 2.1%

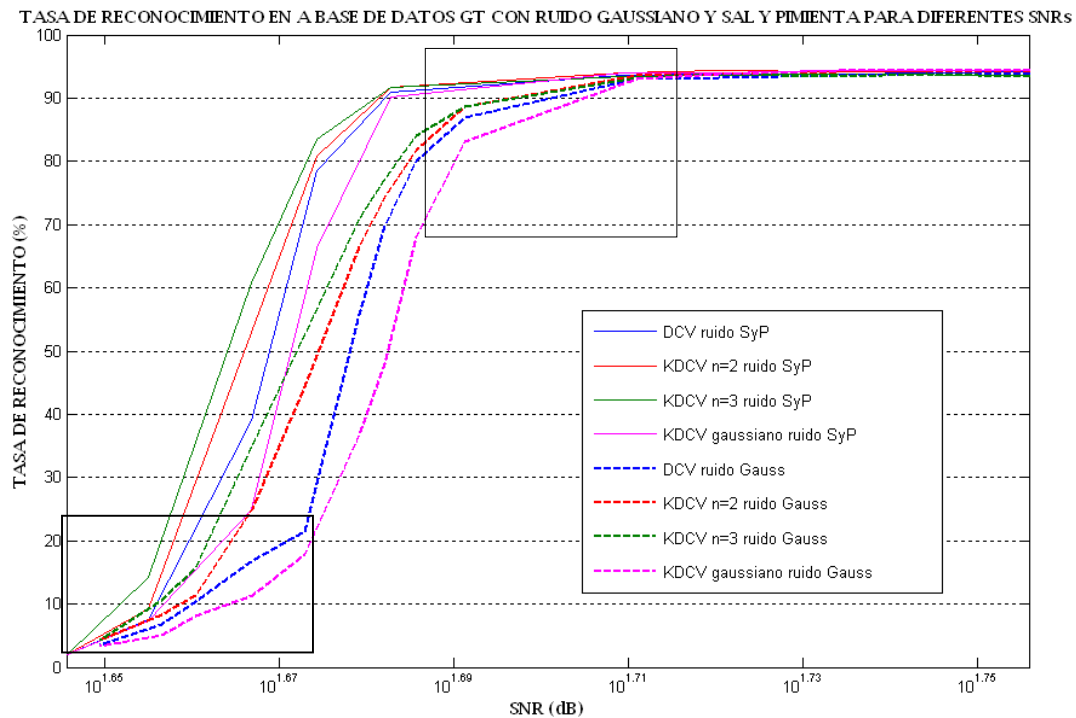
En el tercer tramo se aprecian mayores diferencias. Parece lógico este hecho ya que en las imágenes mostradas anteriormente podemos ver que a partir de un ruido Sal y Pimienta de densidad 40% es bastante difícil distinguir la foto con el ojo humano. Los mejores resultados los ofrece el método KDCV con $n=3$, le sigue el método KDCV con $n=2$, después el método DCV y por último el método KDCV con kernel gaussiano. Este último pasa de tener una tasa de 90.09% para $d=40\%$ a una tasa de 7.43% para $d=80\%$; es decir, ve enormemente reducida su tasa, un 82%. El método DCV reduce la tasa casi un 73% llegando a un 7.63% de aciertos. El siguiente, el método KDCV con $n=2$ empeora los resultados en un 81% llegando a una tasa de éxito de en torno al 9.3%. Finalmente, el más óptimo en este tramo reduce su tasa de un 91.66% a un 14.23%, es decir se reduce casi un 78%. Vemos como en este tramo descienden las tasas y parecen converger a una tasa muy pequeña similar para todos los tramos, manteniéndose ligeramente superior la tasa conseguida por el método KDCV con $n=3$.

Finalmente, en el cuarto tramo todas las tasas convergen a una tasa del 2% para una densidad de ruido sal y pimienta del 100%. Esto es perfectamente lógico, pues para un 100 % de densidad tenemos únicamente ruido en la imagen. Por eso todos los métodos hacen un reconocimiento aleatorio ya que no disponen de información. Este resultado aleatorio corresponde al 2% ya que disponemos de 50 individuos.

Para poder comparar el efecto de los dos tipos de ruido en nuestros métodos de reconocimiento basados en Vectores Comunes Discriminativos y poder contrastarlos, representamos las tasas calculadas con cada método en función de la Relación Señal a Ruido (SNR) que existe. Para ello, calculamos la Potencia de la imagen original y la dividimos entre la potencia de ruido que proporcionan el ruido Gaussiano de una varianza, v , dada; y el ruido Sal y Pimienta para una densidad de ruido, d , dada. A continuación representamos estos resultados en las Gráficas 4.12, 4.13 y 4.14.

GRÁFICA 4.12

TASA DE RECONOCIMIENTO EN LA BASE DE DATOS GT CON PRESENCIA DE RUIDO GAUSSIANO Y RUIDO SAL Y PIMIENTA EN FUNCIÓN DE LA SNR

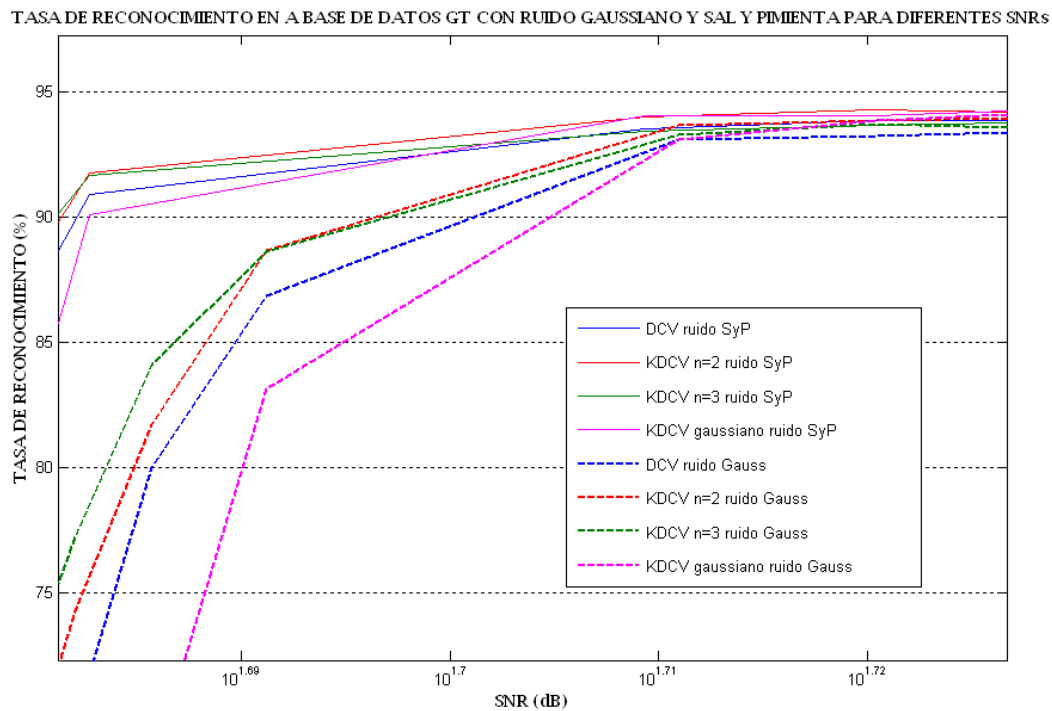


Podemos observar en las anteriores figuras como para una SNR de más de $10^{1.71}$ dB las tasas de reconocimiento para ambos ruidos es aproximadamente la misma, 94.3% y constante para todos los métodos. Si bien las tasas son muy parecidas, existe una pequeña diferencia, los métodos con kernel gaussiano y polinómico de segundo grado en el caso de ruido Sal y Pimienta obtienen unas tasas ligeramente mayores. Como podemos comprobar, el ruido es lo suficientemente bajo como para que no influya en el reconocimiento. Esta SNR (51.4 dB aproximadamente) se corresponde con un ruido

Gaussiano de varianza $v=0.2$ y una densidad de ruido Sal y Pimienta $d=15\%$ aproximadamente. Cuando, a partir de este punto, se decrementa la SNR, también lo hacen las tasas de reconocimiento. Hasta una SNR de 49dB aproximadamente ($10^{1.69}$ dB) en el caso de ruido Gaussiano y una SNR de 48 dB ($10^{1.68}$ dB) en el de Sal y Pimienta, el reconocimiento disminuye suavemente y con más pendiente en el segundo caso. Por tanto, se mantiene una capacidad de generalización cuando lo que tenemos es ruido Sal y Pimienta.

GRÁFICA 4.13

DETALLE DE LA TASA DE RECONOCIMIENTO EN LA BASE DE DATOS GT CON PRESENCIA DE RUIDO GAUSSIANO Y RUIDO SAL Y PIMIENTA EN FUNCIÓN DE LA SNR

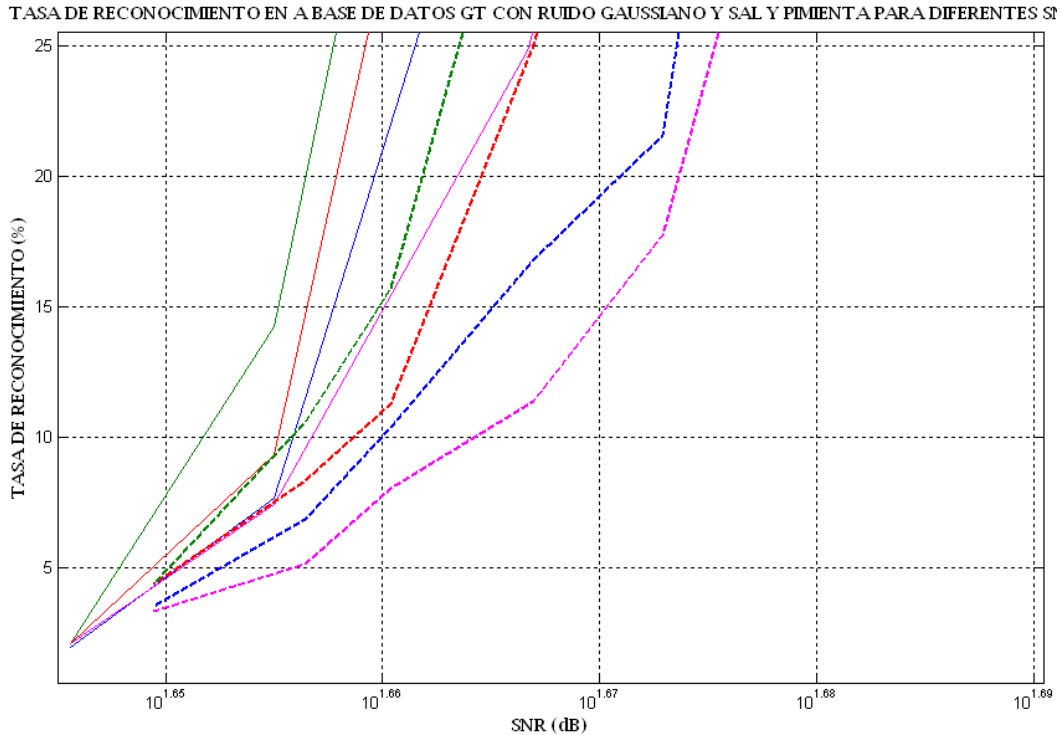


Analizamos entonces el intervalo comprendido entre 51.4 dB y 49 dB de SNR. En el punto donde las tasas empiezan a descender ($10^{1.71}$ dB), el mayor reconocimiento se produce cuando tenemos ruido Sal y Pimienta para los métodos polinómico de $n=2$ y kernel gaussiano, con una tasa del 94%. Después se sitúa el método con kernel polinómico de segundo grado en presencia de ruido Gaussiano con una tasa del 93.7%. Los métodos lineal y polinómico de $n=3$ se sitúan por debajo en el caso de ruido Sal y Pimienta. Por último, se encuentran el resto de métodos con ruido Gaussiano, aunque la diferencia es de, como mucho, un 1%. A partir de este punto se aprecia una clara

tendencia de descenso en las tasas. El método que más empeora es el de kernel gaussiano en ambos casos. Lo hace en un 10.3% en presencia de ruido Gaussiano y casi un 3% en el de Sal y Pimienta; llegando a unas tasas del 83% y 91.2% respectivamente (para una SNR de 49dB). El segundo método que más empeora es el lineal, que hasta este punto disminuye su tasa un 4% en presencia de ruido Gaussiano y un 2% en presencia del ruido Sal y Pimienta. Los métodos con kernel polinómico producen tasas muy similares en este intervalo, llegando a una tasa del 92% con Ruido Sal y Pimienta y 88.6% en el de Gaussiano.

A partir de una SNR de $10^{1.69}$ dB (48.9 dB) comienzan a descender las tasas de manera más abrupta y con más diferencias entre los métodos y los dos ruidos aplicados. Desde este punto y hasta valores de ruido muy bajos (una SNR de 45.7 dB aproximadamente), en general, nuestros métodos generalizan peor en presencia de ruido Gaussiano. El comportamiento global de los métodos para ambos ruidos es similar. Es decir, en ambos casos el que peor se comporta frente al ruido es el método de kernel gaussiano, le sigue el método lineal y, por último, los que mejor se comportan son los de kernel polinómico, siendo el de grado 3 el más óptimo. Otra observación que podemos hacer es que las diferencias entre las técnicas aplicadas son mayores cuando aplico ruido Gaussiano a las imágenes. En este intervalo, se produce un descenso general de aproximadamente un 85%.

Tomamos un punto medio del intervalo, por ejemplo una SNR de $10^{1.67}$ dB para analizar las diferencias entre los resultados obtenidos por los métodos en ambos casos de ruido. En este punto, tenemos una varianza de ruido Gaussiano de aproximadamente 1 y una densidad de ruido Sal y Pimienta del 70%. El reconocimiento es mayor cuando aplicamos el método KDCV de $n=3$, el porcentaje de aciertos es del 70.5% cuando aplico ruido Sal y Pimienta y de 46.8%. Es decir, entre los métodos que mejor generalizan, existe una diferencia muy elevada, de casi 25%. El que peor generaliza es el KDCV con kernel gaussiano y éste obtiene en este punto una tasa del 42.6% con ruido Sal y Pimienta y 14.65%. La diferencia en este caso es del 28%, parecida a la del otro método. Los otros métodos, el KDCV con $n=2$ y el DCV se sitúan con unas tasa del 64.9% y 56% respectivamente en el caso de ruido Gaussiano y 35% y 19.3% en el caso de Ruido Sal y Pimienta. Como podemos apreciar, estas diferencias son algo más grandes, de aproximadamente 30% en el caso lineal y 37% en el polinómico de segundo grado.

GRÁFICA 4.14**DETALLE DE LA TASA DE RECONOCIMIENTO EN LA BASE DE DATOS GT CON PRESENCIA DE RUIDO GAUSSIANO Y RUIDO SAL Y PIMIENTA EN FUNCIÓN DE LA SNR**

Para valores muy grandes de ruido, vemos como las tasas convergen todas ellas a un punto en el que la tasa de reconocimiento es del 2% para una SNR de 44.22 en el caso de ruido Sal y Pimienta. Como ya hemos explicado anteriormente, esto es perfectamente lógico, pues se aproximan a la tasa que conseguiríamos en el peor de los casos, cuando el reconocimiento fuera aleatorio. Sin embargo, en el caso de ruido Gaussiano, no hemos llegado a este valor de SNR, pero también parecen converger los resultados a una tasa del 2%. No obstante, notamos que los valores obtenidos son muy elevados teniendo en cuenta las imágenes con ruido que hemos mostrado. Para el ojo humano, parece imposible reconocer cuando tenemos varianzas de 1 o más y densidades de ruido Sal y Pimienta mayores que 0.6.

4.7. Robustez de los métodos DCV y KDCV frente a la reducción de la resolución

Por último, vamos a estudiar el comportamiento de los métodos basados en vectores comunes discriminativos cuando utilizamos imágenes de tamaño diferente. En este caso usamos la base de datos de imágenes IMM Frontal [62] ya que de las bases de datos estudiadas, es la que contiene las muestras de mayor tamaño. Las imágenes son de 2560×1920 píxeles, en color. El número de individuos en la base de datos es de 12 y disponemos de 10 imágenes por cada uno de ellos.

En principio, pretendemos utilizar las imágenes de la base de datos previamente pasadas a escala de gris y sin ningún tipo de preprocesado adicional. Tras intentar llevar a cabo los algoritmos con los programas implementados en *matlab* nos encontramos con un problema de memoria. Las imágenes son demasiado pesadas y al cargarlas e intentar formar las matrices necesarias para el cálculo de los vectores comunes tenemos serias dificultades. Debido a estas limitaciones, nos vemos obligados a reducir el tamaño de las muestras en 5, de manera que en el posterior análisis partiremos de imágenes de un tamaño de 512×384 píxeles.

Para nuestro experimento, elegimos aleatoriamente 2 imágenes de cada individuo para entrenar cada uno de los métodos y dejamos las 8 restantes para realizar la evaluación. En primer lugar lo hacemos con las imágenes de 512×384 píxeles de las que partimos inicialmente. Repetimos este proceso 10 veces y calculamos las tasas de reconocimiento de cada método como la media de las tasas obtenidas en cada iteración para cada método.

A continuación diezmamos las imágenes, es decir reducimos su tamaño en un factor de 2 (las imágenes adquieren un tamaño de 256×192 píxeles) y procedemos como en el primer caso: entrenamos los métodos con 2 imágenes de cada individuo elegidas al

azar y evaluamos con las 8 restantes. Repetimos el proceso 10 veces con diferentes subconjuntos de entrenamiento y test y calculamos las tasas de reconocimiento de cada método como la media de las tasas obtenidas en cada ejecución.

Después hacemos lo propio con otros factores de diezmado. Lo hacemos para un factor 4 (imágenes de tamaño 128*96 píxeles), 6 (imágenes de tamaño 85*64 píxeles), 8 (imágenes de tamaño 64*48 píxeles), 10 (imágenes de tamaño 51*38 píxeles), 12 (imágenes de tamaño 43*32 píxeles), 15 (imágenes de tamaño 34*26 píxeles), 20 (imágenes de tamaño 26*19 píxeles), 30 (imágenes de tamaño 17*13 píxeles) y 50(imágenes de tamaño 10*8 píxeles).

En principio pensábamos hacer como habíamos hecho con factores anteriores, es decir, entrenar los métodos con las imágenes iniciales y evaluarlos con las muestras en las que habíamos introducido las variaciones como en el caso de ruido, oclusiones o luz lateral. En este caso, no podíamos entrenar con un tamaño de muestra y evaluar con otro ya que estos métodos se basan en la información que da cada píxel y por tanto para realizar el reconocimiento es necesario que en todo momento mantengamos el tamaño de las muestras constante. Por eso, aunque vamos reduciendo el tamaño en cada experimento, mantenemos el mismo tanto para entrenamiento como para test.

A continuación, en la Tabla 4.11 mostramos los resultados obtenidos y los representamos en la Gráfica 4.15

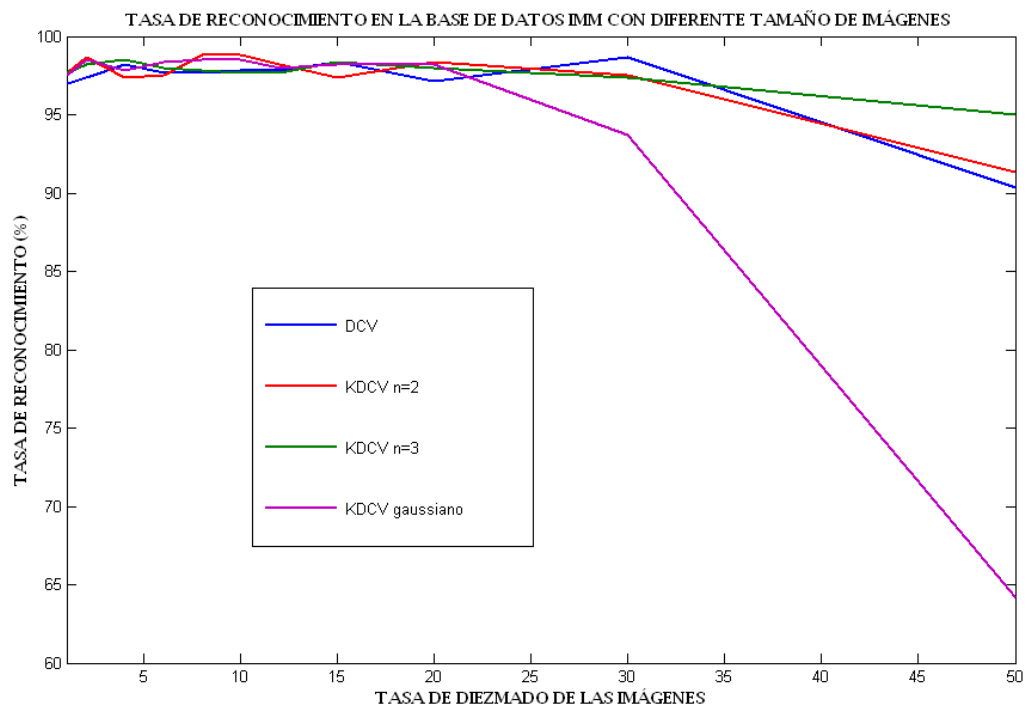
TABLA 4.11

TASA DE RECONOCIMIENTO PAR LA BASE DE DATOS IMM CON REDUCCIÓN DE LA DIMENSIÓN

Tamaño de las imágenes (píxeles)	Tasa de diezmado	Tasas de Reconocimiento (%) y Desviaciones Típicas							
		DCV		Kernel DCV n=2		Kernel DCV n=3		Kernel Gaussiano	
		tasa%	Σ	tasa%	σ	tasa%	σ	tasa%	σ
512 *384	x = 1	97.00	2.33	97.67	2.33	97.67	1.72	97.50	1.96
256*192	x = 2	97.33	1.96	98.67	1.31	98.17	1.46	98.50	1.23
128*96	x = 4	98.17	1.46	97.33	1.61	98.50	2.77	97.83	1.12
85*64	x = 6	97.67	2.38	97.50	1.18	98.00	1.96	98.33	1.57
64*48	x = 8	97.67	1.16	98.83	1.12	97.83	1.12	98.50	0.53
51*38	x = 10	97.83	0.80	98.83	1.37	97.67	2.25	98.50	1.23
43*32	x = 12	97.83	1.76	98.17	0.95	97.67	1.23	98.00	1.53
34*26	x = 15	98.33	1.11	97.33	2.85	98.33	1.36	98.17	1.23
26*19	x = 20	97.17	2.50	98.33	1.76	98.00	1.32	98.17	1.66
17*13	x = 30	98.67	1.31	97.50	1.62	97.33	2.38	93.67	4.07
10*8	x = 50	90.33	3.67	91.33	3.91	95.00	2.36	64.17	10.46

GRÁFICA 4.15

TASA DE RECONOCIMIENTO EN LA BASE DE DATOS IMM CON DIFERENTES TAMAÑOS DE LAS MUESTRAS



Como podemos observar, no conseguimos obtener unas tasas que realmente se diferencien en función del diezmado aplicado a las imágenes, hasta que no tenemos un factor de 25 o superior. Este factor de 25 se corresponde con un tamaño de las imágenes de 15×20 píxeles. Si recordamos que partimos de imágenes previamente redimensionadas de las originales (que tenían un tamaño de 2560×1920 píxeles), este factor es, realmente, 125. Se obtienen muy buenas tasas, entre 97% y 99%, para todos los métodos desde que tenemos imágenes de 512×384 píxeles hasta 15×20 píxeles. Los métodos van alternando el valor de sus tasas en este intervalo sin una clara tendencia. Todos ellos, incrementan ligeramente su rendimiento para algunos factores, que ayudan a centrar las imágenes. No hay un claro dominio de un método respecto a los demás cuando tenemos tamaños elevados.

A partir de un diezmado de 20, es decir, un tamaño de imagen de 26×19 píxeles, observamos que el método KDCV con kernel gaussiano, comienza a empeorar. Empeora linealmente hasta que tenemos un tamaño de imagen de 17×13 píxeles para luego hacerlo de manera más abrupta. Pasa de proporcionar el 98.17% de aciertos al 64.17% en el caso de un tamaño de imagen muy pequeño como es 10×8 píxeles.

En el caso del método DCV lineal se observa un comportamiento muy parecido. La diferencia es que antes de comenzar su empeoramiento, ve incrementada su tasa de reconocimiento ligeramente en el caso de un factor de diezmado 30, en el que el éxito es del 98.67%. Después empeora pero de manera más suave hasta obtener una tasa del 90.33% para un tamaño de muestras de 10×8 píxeles. Por tanto, ve decrementada su tasa en un 8.34%.

Para los métodos con kernel polinómico, el comportamiento es más estable. Hasta que aplicamos un factor de diezmado de 30, su actuación apenas difiere y mantienen menos cambios que en los casos anteriores. Se mantiene aún más estable, si cabe, el método KDCV con $n=3$. Después, a partir de un tamaño de imagen de 17×13 píxeles comienzan un ligero empeoramiento, más suave en el caso del kernel polinómico de tercer grado. Este proporciona una tasa del 95% cuando evaluamos con el máximo factor de diezmado aplicado en el experimento, mientras que el KDCV de $n=2$ produce una tasa de reconocimiento del 91.33%. Como hemos podido observar, estos métodos mantienen su capacidad de generalización disminuyendo su porcentaje de aciertos tan

sólo un 3.5% el de grado 3 y un 7.5% el de grado 2 desde que obtienen su mayor tasa de aciertos.

Capítulo:

5. Conclusiones

En este Capítulo vamos a hacer un análisis y reflexión sobre la robustez de los métodos de reconocimiento basados en Vectores Comunes Discriminativos a la vista de los resultados obtenidos en el apartado de experimentos. De manera que nos vamos a fijar en el comportamiento que han tenido los métodos DCV lineal, y KDCV con núcleos polinómicos de segundo y tercer grado y el método KDCV con núcleo gaussiano.

Lo cierto es que el coste de entrenamiento, la velocidad de ejecución, y las necesidades de almacenamiento son algunos factores que pueden ser utilizados para evaluar un método de reconocimiento facial. Aquí, las diferencias entre estos factores no van a ser discutidas. En este apartado nuestro objetivo es analizar si los métodos estudiados alcanzan un comportamiento robusto frente a diferentes tipos de variabilidad de los datos y en consecuencia si consiguen un buen poder de generalización. Los factores a los que vamos a someter estos métodos van a ser: el número de individuos en la base de datos; el número de muestras con las que entrenamos nuestros métodos; la introducción de iluminación lateral, la presencia de oclusiones en las imágenes, la existencia de ruido Gaussiano y ruido Sal y Pimienta en las muestras, o el

redimensionamiento de las mismas. Queremos estudiar la capacidad de generalización de los métodos presentados basados en vectores comunes, unos respecto de otros. Lo hacemos frente a un determinado rango de problemas con cierta variabilidad.

1. En este proyecto, hemos obtenido vectores de proyección óptima en el espacio nulo de la dispersión dentro de la clase. Hemos comprobado que cada muestra en una clase dada produce el mismo y único vector común cuando son proyectadas sobre el espacio nulo de S_W . Los vectores de proyección óptima se encuentran usando los vectores comunes y los vectores comunes discriminativos son determinados mediante la proyección de cualquier muestra de cada clase sobre el espacio de los vectores de proyección. No hay pérdidas de información contenidas en nuestros métodos, en el sentido de que los métodos tienen el 100 % de la tasa de reconocimiento para el conjunto de entrenamiento.

2. Lo primero que podemos destacar de los experimentos realizados, es el hecho de que todos los métodos, cuando los evaluamos en una situación en la que ningún factor cambia, producen unas tasas muy similares; es decir, en general no hay una diferencia significativa entre las tasas de reconocimiento de los métodos lineal y kernel para la base de datos de caras, ya que las muestras de caras son linealmente separables. Esto puede atribuirse a la naturaleza de las imágenes de las caras en la base de datos. Las imágenes de los individuos son en la mayoría de posición frontal, y las condiciones lumínicas son similares. Por tanto, las caras en la base de datos son linealmente separables. En tales casos, utilizando las correlaciones de orden superior a través de los núcleos puede degradar el comportamiento como en nuestro caso ya que el problema está cerca de ser linealmente separable. Las aproximaciones no lineales son aplicadas cuando las proyecciones lineales no son suficientes para representar los datos de manera que permitan la extracción de características discriminatorias. Estas aproximaciones no lineales tratan de proyectar los datos en un espacio de mayor dimensión donde las clases sean más propensas a ser linealmente separables. Este es el caso para distribuciones más complejas donde meros hiper-planos no fallen al tratar de separar las clases a distinguir.

Podemos afirmar entonces que las caras son datos que están cerca de ser linealmente separables.

En las tareas de procesamiento de señales, frecuentemente hay mucha redundancia en las respectivas imágenes/vectores porque, en primer lugar, las imágenes de algunas clases son bastante similares y, en segundo lugar, los píxeles próximos en una imagen están altamente correlados. Sin embargo, el método kernel DCV podría mejorar los resultados del reconocimiento del método DCV lineal en diferentes bases de datos de caras que tengan distribuciones complejas y no lineales.

3. En nuestro primer caso experimental, hemos evaluado el método DCV en dos bases de datos, YALE y AR, de diferente tamaño. La primera base de datos de imágenes faciales está formada por un menor número de sujetos y también de muestras por sujeto. El método DCV extrae las propiedades comunes de las clases en el conjunto de entrenamiento mediante la eliminación de las diferencias de las muestras en cada clase. Hemos podido observar cómo las tasas son superiores cuando usamos la base de datos AR. Observamos una diferencia importante en los resultados, llegando esta incluso a más de un 10% cuando empleamos el método lineal. La tasa de reconocimiento para la base de datos Yale con el método DCV ha sido significativamente más baja. Las imágenes sin clasificar para este método fueron típicamente imágenes que no se tomaron bajo las condiciones de ambiente estándar que se usaron para la mayoría de los datos (es decir, iluminación centrada, izquierda o derecha). Dadas las direcciones de proyección encontradas por el método, se espera que este método funcione bien cuando las muestras de test de un sujeto son similares a las muestras del sujeto tomadas para entrenamiento. Dado que el método “dejar uno fuera” fue utilizado para test y que hubo solamente una muestra de estas condiciones de iluminación por clase, estas imágenes de iluminación inusual se comportan como datos anómalos (es decir, estas imágenes distan mucho de las muestras utilizadas para el entrenamiento). Serían de esperar mejores resultados si hubiera más de una muestra con estas condiciones de iluminación.

4. Reforzamos las conclusiones anteriores a cerca de la influencia de la luz en el reconocimiento, cuando comprobamos la robustez de los métodos cuando nos enfrentamos a cambios en la iluminación en el apartado experimental 4.4.

Comprobamos cómo cuando se evalúan los métodos con imágenes similares a las utilizadas para entrenar, es decir, sin cambios de iluminación, las tasas que se obtienen son muy elevadas y similares para todos los métodos. En el momento en el que queremos evaluar con datos anómalos, es decir, imágenes con iluminación lateral, los métodos reducen bastante sus prestaciones. Reducen sus tasas en más de un 50%. No obstante, el método lineal es el que menos se degrada, (algo más de la mitad) junto con el método KDCV con kernel gaussiano. Los métodos con kernel polinómico son los más sensibles a esta variación, quedando las tasas a $1/3$ en el caso de segundo grado y a $1/4$ en el caso de grado tres, aproximadamente.

5. Tras la realización del experimento mencionado en el apartado 4.2, en el que variamos el número de individuos que componen la base de datos; comprobamos que el éxito de los métodos tiene una alta dependencia con el tamaño del espacio nulo de la matriz de dispersión intra-clase, S_W , ($N(S_W)$). De $N(S_W)$ se extraen los vectores comunes que representan a cada persona en el conjunto de entrenamiento. El espacio nulo es el que generan los autovectores correspondientes a autovalores nulos de la matriz S_W . Cuando C es mayor, como se explica en [51], el tamaño del espacio nulo disminuye y se dispone de menos información para obtener los vectores de proyección óptimos.

Comprobamos este hecho cuando vamos aumentando el número de individuos en la base de datos ORL. Observamos como el rendimiento decrece con la reducción de la dimensión del espacio nulo, es decir, cuando aumenta el número de sujetos. Inicialmente, cuando el conjunto es pequeño observamos tasas de reconocimiento muy elevadas y similares para todos los métodos, siendo del 100% en el caso lineal. Como es de esperar, es más fácil distinguir a una persona entre un conjunto pequeño de ellas. A medida que aumenta C , los métodos van empeorando pero sus tasas no se ven reducidas demasiado, siguen siendo altas para un conjunto de 40 personas. El más robusto frente a este factor, resulta ser el método KDCV con kernel Gaussiano. A continuación se sitúan los demás, que disminuyen su rendimiento en una ligera mayor proporción y aproximadamente de la misma magnitud. Aunque ofrece un ligero menor rendimiento el método KDCV con $n=3$, siendo el comportamiento de los métodos lineal y de kernel polinómico de segundo grado, equivalentes.

6. A continuación, en el apartado experimental 4.3, evaluamos el impacto producido por la elección del número de muestras con las que entrenamos nuestros algoritmos. En la práctica, es interesante conocer el comportamiento de los métodos con un número reducido de muestras de entrenamiento ya que, obviamente, todos los métodos se aproximan al óptimo cuando este tamaño crece.

Los métodos aumentan su rendimiento cuando se entrenan con un mayor número de muestras. Para un número bajo de muestras de entrenamiento, el método más eficiente es el de kernel Gaussiano. Los métodos lineal y polinómico de grado 2 son equivalentes y por último, el de grado 3 es el más pobre. No obstante, para 3 muestras de entrenamiento las tasas obtenidas son altas, siendo en casi todos los casos superiores al 90%. Cuando aumentamos el número de muestras de entrenamiento, las diferencias entre los métodos se reducen, consiguiendo, todos, tasas más elevadas y más cercanas aunque se mantiene el ranking de rendimiento. Es decir, sigue siendo el método KDCV con kernel gaussiano el que mayor capacidad de generalización tiene y el de kernel polinómico de grado 3 el que menos.

7. Cuando combinamos las dos variables anteriores, cambiamos el número de sujetos y el número de muestras de entrenamiento; concluimos que las tasas de reconocimiento decrecen en menor medida cuando disminuimos el espacio nulo de S_W en el caso en el que entrenamos con más número de imágenes.

8. El impacto de las oclusiones también ha sido evaluado. Cuando entrenamos y evaluamos el reconocimiento sin presencia de oclusiones, las tasas obtenidas son muy elevadas para todos los métodos. Cuando introducimos oclusiones en las imágenes, el rendimiento de todos los métodos empeora en gran medida. Se reduce la capacidad de generalización en todos los casos. Los métodos que más robustos permanecen frente a este factor son los métodos KDCV con kernel polinómico, y aún un poco más, el de tercer grado. El método que más se degrada es el lineal, haciéndolo en un 45%.

Podemos afirmar que existe una debilidad en esta técnica de reconocimiento facial basada en vectores comunes: el alto grado de sensibilidad frente a oclusiones parciales y a cambios en la iluminación. Como los métodos hacen un análisis basado en la

información que proporciona cada píxel, cuándo a cierta cantidad de píxeles se les priva de información, es lógico que estos métodos ofrezcan peores resultados.

9. Respecto al ruido, podemos afirmar que los métodos basados en Vectores Comunes Discriminativos son muy robustos. Lo son más, si cabe, respecto al ruido Sal y Pimienta que respecto a ruido Gaussiano. Observamos que para imágenes con gran cantidad de ruido, los métodos siguen realizando un reconocimiento bastante óptimo. De entre los métodos estudiados, el que mejor generaliza en presencia de ambos tipos de ruido es el KDCV con kernel polinómico de $n=3$. Éste produce las tasas más altas, seguido del de $n=2$. El menos robusto es el KDCV con kernel gaussiano, que es más sensible a la presencia de ruido en las imágenes.

En el peor de los casos, cuando el ruido es total, estos métodos ofrecen un resultado igual al producido por la elección aleatoria, pero incluso con ruido muy elevado ofrecen unas tasas relativamente altas si lo comparamos con el reconocimiento que haría un humano. Esto puede ser debido a que la base de datos utilizada contiene unas imágenes con gran cantidad de información aparte de la cara de los sujetos. Tienen información de fondo y vestimenta que pueden ayudar al reconocimiento. Aunque el fondo es común para todos los individuos, la vestimenta no lo es, y puede ayudar al reconocimiento de los individuos ya que este tipo de métodos utiliza la información de cada píxel en la imagen y los trata todos como si fueran parte de la “cara”.

10. Por último, estos métodos son muy robustos frente al diezmado de las imágenes. Producen tasas muy elevadas, todos ellos, incluso hasta un tamaño de imagen de 20×15 píxeles. A partir de una mínima resolución, como pueden ser 20×15 píxeles, podríamos pensar que el rendimiento aumente a medida que lo hace el tamaño de las imágenes pero esto no es así. Se consiguen tasas similares, siendo incluso, algunas tasas, superiores a otras producidas con más resolución. Esto se debe a que a veces, el diezmar la imagen implica centrarla y los métodos, que toman la información de cada píxel, son más eficientes. Si por ejemplo, tenemos un ojo situado en cuatro píxeles, un ligero movimiento de una imagen a otra podría ofrecer peores resultado que si lo tengo centrado en un único píxel.

A partir de este tamaño indicado, las tasas descienden en todos los métodos, siendo el de kernel gaussiano el más sensible a la compresión. El método que más robusto se mantiene es el KDCV con $n=3$, que apenas es sensible. Después se encuentran los métodos lineal y polinómico de grado 2, que son ligeramente más sensibles, pero no tanto como lo es el de kernel gaussiano.

5.1. Comparación del comportamiento de los métodos

A continuación, mostramos un ranking de robustez de los métodos basados en Vectores Comunes Discriminativos respecto a las variabilidades analizadas anteriormente.

TABLA 5.1
COMPARACIÓN DEL COMPORTAMIENTO DE LOS MÉTODOS RESPECTO A LOS FACTORES ESTUDIADOS

Ranking Comp.	# Individuos	# N train⁽²⁾	Iluminación n	Oclusiones	Ruido	Diezmado
1	KDCV Gaussiano	KDCV Gaussiano	DCV lineal	KDCV n=3	KDCV n=3	KDCV n=3
2	DCV lineal ⁽¹⁾	DCV lineal	KDCV Gaussiano	KDCV n=2	KDCV n=2	KDCV n=2
3	KDCV n=2 ⁽¹⁾	KDCV n=2	KDCV n=2	KDCV Gaussiano	DCV lineal	DCV lineal
4	KDCV n=3	KDCV n=3	KDCV n=3	DCV lineal	KDCV Gaussiano	KDCV Gaussiano

(1) Ambos métodos ofrecen muy similares tasas de rendimiento

(2) Comportamiento respecto a la reducción del número de muestras de entrenamiento.

El método KDCV con kernel gaussiano es el más robusto frente al número de individuos en la base de datos así como a la reducción del número de imágenes de entrenamiento. Sin embargo, es el más sensible frente a ruido y diezmado de las muestras.

El KDCV con kernel polinómico de $n=3$ es el método más consistente si introducimos oclusiones parciales o ruido en las imágenes pero es el más débil frente a la iluminación parcial, el aumento del número de individuos de la base de datos así como la disminución del número de muestras que usamos para entrenar.

El método DCV lineal es el más estable frente a la introducción de iluminación parcial en la evaluación, mientras que muestra la mayor sensibilidad frente a oclusiones.

Por último, el método KDCV con kernel polinómico de $n=2$ es el que ofrece un comportamiento menos variable respecto a los otros métodos, manteniéndose frente a todos los factores entre una posición segunda y tercera.

Referencias

- [1] http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4127500
- [2] W. Zhao, Sarnoff Corporation (National Institute of Standards and Technology), “Face Recognition: A Literature Survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 35, Issue 4, Dec. 2003
- [3] R. Chellappa, C.L. Wilson, and S. Sirohey, “Human and Machine Recognition of Faces: A Survey,” Proc. IEEE, vol. 83, pp. 705-740, May 1995
- [4] I. Biederman and P. Kalocsai. Neurocomputational bases of object and face recognition. Philosophical Transactions: Biological Sciences, 352(1358): 1203–1219, 1997.
- [5] I. Gauthier and M. Behrmann. Can face recognition really be dissociated from object recognition? Journal of Cognitive Neuroscience, 11(4), 1999.
- [6] C. Darwin. The Expression of the Emotions in Man and Animals. London: John Murray, 1872.

- [7] F. Galton. Personal identification and description. *Nature*, pages 173–188, 1888.
- [8] M. A. Fischler and R. A. Elschlager. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 22(1):67–92, 1973.
- [9] A. Yuille, P. Hallinan, and D. Cohen. Feature extraction from faces using deformable templates. *International Journal of computer Vision*, 1992.
- [10] M. Turk and A. P. Pentland. Face recognition using eigenfaces. *IEEE Conf. Computer Vision and Pattern Recognition*, 1991.
- [11] S. Chen and D. Li. Modified linear discriminant analysis. *Pattern Recognition*, pages 441–443, 2005.
- [12] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: *Recognition using class specific linear projection*. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.
- [13] M. H. Yang. Kernel Eigenfaces vs. Kernel Fisherfaces: Face recognition using kernel methods. *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 205–211, 2002.
- [14] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using Laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005.
- [15] “Introduction to aspects of face processing: Ten questions in need of answers,” in *Aspects of Face Processing*, H. Ellis, M. Jeeves, F. Newcombe,
- [16] A. P. Ginsburg, “*Visual information processing based on spatial filters constrained by biological data*,” AMRL Tech. Rep., pp.78-129, 1978.
- [17] L. D. Harmon, “The recognition of faces,” *Scientific American*, vol. 229, pp. 71-82, 1973.

- [18] B. A. Golomb and T. J. Sejnowski, "SEXNET: A neural network identifies sex from human faces," in *Advances in Neural Information Processing Systems* 3, D. S. Touretzky and R. Lipmann, Eds. San Mateo, CA: Morgan Kaufmann, 1991, pp 572-577
- [19] A. G. Goldstein, "Facial feature variation: Anthropometric data 11," *Bull. Psychonomic Soc.*, vol. 13, pp. 191-193, 1979.
- [20] F. Rosenblatt. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan, Washington, DC, 1962.
- [21] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition*, pages 586–591, 1991.
- [22] R. A. Fisher. The use of multiple measures in taxonomic problems. *Annals of Eugenics*, 17:179–188, 1936
- [23] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegmann. *Eigenfaces vs Fisherfaces: Recognition using class specific linear projection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(7):711–720, 1997.
- [24] A. Bell and T. Sejnowski. *An information maximization approach to blind separation and blind deconvolution*. Neural Computation, 7:1129–1159, 1995.
- [25] C. Jutten and J. Herault. Blind separation of sources. *Signal Processing*, 24:1–10, 1991
- [26] J. F. Cardoso. *Higher-order contrasts for independent component analysis*. Neural Computation, 11(1):157–192, 1999.
- [27] A. Hyvärinen and E. Oja. *A fast fixed-point algorithm for independent component analysis*. Neural Computation, 9(7):1483–1492, 1997.
- [28] J. Yang, D. Zhang, and A. F. Frangi. *Two-dimensional PCA: A new approach to appearance-based face representation and recognition*. IEEE

Transactions on Pattern Analysis and Machine Intelligence, 26(1):131– 137, 2004.

- [29] M. Visani, C. Garcia, and J. M. Jolion. *Two-dimensional-oriented linear discriminant analysis for face recognition*. In Proceedings of the International Conference on Computer Vision and Graphics, pages 1008–1017, 2004.
- [30] T. Cootes, G. Edwards, and C. Taylor. *Active appearance models*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(6):681–685, 2001.
- [31] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. *Active shape models—their training and application*. Computer Vision Graphics and Image Understanding, 61(1):38–59, 1995.
- [32] L. R. Rabiner. *A tutorial on hidden markov models and selected applications in speech recognition*. Proceedings of the IEEE, 77(2):257–285, 1989.
- [33] L. R. Rabiner and Juang B.-H. *Fundamentals of Speech Recognition*. Prentice Hall, Eaglewood Cliffs, NJ, 1993.
- [34] Yoav Freund and Robert E. Schapire. *A decision-theoretic generalization of on-line learning and an application to boosting*. In Second International Conference on Computational Learning Theory, 1995.
- [35] B. E. Boser, I. M. Guyon, and V. Vapnik. *A training algorithm for optimal margin classifiers*. In Fifth Annual Workshop on Computational Learning Theory, pages 144–152, Pittsburgh ACM, 1992.
- [36] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.
- [37] C. Cortes and V. Vapnik. *Support vector networks*. Machine Learning, 20:1–25, 1995.

- [38] F. H. C. Tivive and A. Bouzerdoum. *A shunting inhibitory convolutional neural network for gender classification*. In Proceedings of the International Conference on Pattern Recognition, 2006.
- [39] G. Csurka, C. Bray, C. Dance, and L. Fan. *Visual categorization with bags of keypoints*. In ECCV Workshop on Statistical Learning in Computer Vision, pages 327–334, 2004.
- [40] F. Rosenblatt. *The perceptron: a probabilistic model for information storage and retrieval in the brain*. Psychological Review, 65:386–408, 1958.
- [41] <http://www.face-rec.org/>
- [42] M. Turk and A.P. Pentland, “Eigenfaces for Recognition,” J. Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.
- [43] W. Zhao, R. Chellapa, and A Krishnaswamy, “Discriminant Analysis of Principal Components for Face Recognition,” Proc. Third IEEE Int’l Conf. Automatic Face and Gesture Recognition, pp. 336-341, Apr. 1998.
- [44] W. Zhao, R. Chellapa, A Rosenfeld, and Phillips, “Face Recognition: A Literature Survey”, Technical Report CAR-TR-948, Univ. of Maryland, College Park, 2000.
- [45] Q. Tian, M. Barbero, Z.H. Gu, and S.H. Lee, “Image Classification by the Foley-Sammon Transform”, Optical Eng., vol. 25, no, 7, pp. 834-840, 1986.
- [46] Y.Q. Cheng, Y.M. Zhuang, and J.Y. Yang, “Optimal Fisher Discriminate Analysis Using the Rank Decomposition”, *Pattern Recognition*, vol. 25, pp. 101-111, 1992.
- [47] D. L. Sweets and J. Weng, “Using Discriminant Eigenfeatures for Image Retrieval”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 831-836, Aug. 1996.
- [48] L.-F. Chen, H.-Y.M. Liao, M.-T. Ko, J.-C. Lin, and G.-J. Yu, “A New LDA-Based Face Recognition System Which Can Solve the Small Sample Size Problem”, *Pattern Recognition*, vol. 33, pp. 1713-1726, 2000.

- [49] R. Huang, Q. Liu, H. Lu, and S. Ma, “Solving the Small Size Problem of LDA”, *Proc. 16th Int’l Conf. Pattern Recognition*, vol. 3, pp. 29-32, Aug. 2002.
- [50] H. Yu and J. Yang, “A Direct LDA Algorithm for High-Dimensional Data with Application to Face Recognition”, *Pattern Recognition*, vol. 34, pp. 2067-2070, 2001.
- [51] Hakan Cevikalp, Marian Neamtu, Mitch Wilkes, and Atalay Barkana. “Discriminative Common Vectors for Face Recognition”, *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, NO. 1, January 2005.
- [52] Hakan Cevikalp, Marian Neamtu, and Mitch Wilkes. “Discriminative Common Vectors Method Whit Kernels”. *IEEE Transactions on Neural Networks*, Vol. 17, No. 6, November 2006.
- [53] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R. Müller, “Fisher discriminant analysis with kernels”, in *Neural Networks for Signal Processing IX*, Y.-H. Hu, J. Larsen, E. Wilson, and S. Douglas, Eds. Piscataway, NJ:IEEE, 1999, pp. 41-48.
- [54] G. Baudat and F. Anouar, “Generalized discriminant analysis using a kernel approach”, *Neural Comput.* Vol. 12, pp. 2385-2404, 2000.
- [55] J. Yang, A. F. Frangi, Z. Jin, and J.-Y. Yang, “Essence of kernel Fiser discriminant: KPCA plus LDA”, *Pattern Recognit.*, vol. 37, pp. 2097-2100, Oct. 2004
- [56] B. Schölkopf, “Support Vector Learning”, Ph.D. dissertation, Informatik der Technischen Universität, Berlin, Germany, 1997.
- [57] K. Tsuda, “*Subspace classifier in reproducing kernel Hilbert space*,” in *Proc. Int. Joint Conf. Neural Netw.*, 1999, vol. 5, pp. 3454–3457.
- [58] <http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html>

- [59] Intelligence, vol. 18, no. 8, pp. 831-836, Aug. 1996. [18] A.M. Martinez and R. Benavente, “*The AR Face Database*,” CVC Technical Report #24, Computer Vision Center (CVC), Barcelona, Spain 1998
- [60] AT&T Laboratories Cambridge, The ORL Database of Faces [Online]. Available: <http://www.uk.research.att.com/facedatabase.html>
- [61] http://www.anefian.com/research/face_reco.htm
- [62] http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=3943
- [63] K Fukunaga, Introduction to Statistical Pattern Recognition, second ed. New York: Academic Press, pp. 31-34, 39-40, 220-221, 1990
- [64] F. Perez-Cruz and O. Bousquet, “Kernel methods and their potential use in signal processing,” *IEEE Signal Process. Mag.*, vol. 21, no. 3, pp. 57–65, May 2004.
- <http://www.biometria.gov.ar/index.php/documentos/92-reconocimiento-facial>
- http://arantxa.ii.uam.es/~msanchez/docencia/maaa/transparencias/Transf_0809.pdf
- <http://www.uv.es/marjoari/pdf/pres.pdf>
- www.ehu.es/ccwintco/uploads/1/13/Dev.pdf
- http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/4014/pdf/imm4014.pdf
- http://www.freidok.uni-freiburg.de/volltexte/4835/pdf/thesis_final.pdf
- <http://cobweb.ecn.purdue.edu/~aleix/fr.htm>
- <http://www.utp.edu.co/php/revistas/ScientiaEtTechnica/docsFTP/14513667-72.pdf>
- http://catarina.udlap.mx/u_dl_a/tales/documentos/msp/gutierrez_g_l/capitulo1.pdf
- <http://intart.wetpaint.com/page/Novedades+en+Redes+Neuronales>

- <http://erevista2.aepia.org/index.php/ia/article/viewFile/1020/1017>

